DOCUMENT RESUME

ED 355 195 SP 034 281

AUTHOR	Izu, Jo Ann; And Others
TITLE	Assessment Component of the California New Teacher Project: Evaluation of Existing Teacher Assessment Practices.
INSTITUTION	Far West Lab. for Educational Research and Development, San Francisco, Calif.
SPONS AGENCY	California Commission of Teacher Credentialing, Sacramento.; California State Dept. of Education, Sacramento.
PUB DATE	Feb 92
CONTRACT	TCC-C021
NOTE	246p.; For related documents, see ED 323 197, ED 342 761, and SP 034 278-282.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC10 Plus Postage.
DESCRIPTORS	Beginning Teachers; *Board of Education Policy; *College Outcomes Assessment; Elementary Secondary Education; Evaluation Criteria; Evaluation Methods; Evaluation Research; Higher Education; Inservice Teacher Education; *Needs Assessment; Policy Formation; Preservice Teacher Education; Research Design; *State Standards; Teacher Certification; *Teacher Evaluation
IDENTIFIERS	*Assessments of Performance in Teaching; *California New Teacher Project; Knowledge Base for Teaching; Reform Efforts

ABSTRACT

This document reports on the third year of the assessment component of the California New Teacher Project (CNTP). In the first 2 years, research focused on pilot testing innovative forms of new teacher assessment. Part of the third year study was a careful review of the state of assessment practices in California. The study was designed to yield information about existing assessment practices at the college and university level, the local school district level, and the State level within the context of the legal, social, and economic situation facing California educational systems and to provide information that would lead to the conceptual base and design of a new system of support and assessment for beginning teachers in California. This report, organized into six chapters, begins with an introduction describing the assessment component of the CNTP, the current context for assessment of California teachers, the essential design features of the study, guiding constructs for an ideal statewide assessment system, and developing interview protocols. The remainder of the report is organized into chapters covering institutions of higher education assessment practices, local education agency assessment practices, the State assessment practices substudy, experiences of beginning teachers, and a summary of findings across substudies and conclusions. Two appendices are attached: Sample Items from the California Basic Educational Skills Test (CBEST); and Comparisons with Phase II Assessments. (LL)

BEST COPY AVAILABLE

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

T. Ross

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

SB 34281

This report is based upon work performed pursuant to Contract TCC-C021 with the California Interagency Task Force of the California Commission on Teacher Credentialing and the California State Department of Education New Teacher Project. The contents do not necessarily reflect the views or policies of these agencies and the accuracy of the information is the sole responsibility of the Far West Laboratory.

Assessment Component of the California New Teacher Project:

Evaluation of Existing Teacher Assessment Practices

Jo Ann Izu
Claudia Long
Kendyll Stansbury
Dennis Tierney

February, 1992

Gary Estes, Project Director
Far West Laboratory for
Educational Research and Development
San Francisco, California

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION 1.1

Assessment Component of the CNTP: Phase II 1.2

Current Context for Assessment of California Teachers 1.4

Essential Design Features of this Study 1.7

Guiding Constructs for an Ideal Statewide Assessment System 1.7

Development of Interview Protocols 1.10

CHAPTER 2: INSTITUTIONS OF HIGHER EDUCATION ASSESSMENT PRACTICES 2.1

Description of Sample and Methods 2.1

Sample Description 2.1
Methodology 2.3

Interview protocols 2.3
Documents collected 2.5
Interview teams 2.9

Institutions of Higher Education 2.10

Overview of Major Assessment Points 2.12

Admission to Credential Program 2.12

Range of methods used 2.12
Purposes and uses of admissions assessments 2.17
Technical quality of admissions assessments 2.20

Retention in Credential Programs 2.21

Range of methods used 2.21
Purposes and uses of retention assessments 2.22
Technical quality of assessments 2.23

Formal Student Teaching 2.24

Range of methods used 2.24
Purposes and uses of assessments 2.25
Technical quality 2.26

Recommendation for the Credential 2.29

Range of methods used 2.29
Purposes and uses of assessments 2.29
Technical quality 2.29

Post Credential Support 2.30

Range of methods used 2.30
Purposes and uses of assessments 2.30
Technical quality 2.30

Focus of IHE Assessments 2.30

Subject-matter 2.31
General Pedagogy 2.31
Content Pedagogy 2.32
Other Constructs 2.32

Basic skills proficiency 2.32
Personal attributes 2.33
Programmatic needs 2.33

Differential Impact of Assessment Practices 2.34

Costs of IHE Assessments 2.37

Admission Costs 2.37
Contributions of the IHE Assessments to Credentialing Recommendations 2.39

Summary and Conclusions 2.40

Key Admissions Elements 2.40
Key Retention and Recommendation Elements 2.41

CHAPTER 3: LOCAL EDUCATION AGENCY ASSESSMENT PRACTICES 3.1

Description of Sample and Methods 3.1

Sample Description 3.1
Methodology 3.4
Interview Protocols 3.4
Documents Collected 3.6
Interviewers 3.6

LEA Context 3.8

Overview of Major Assessment Points 3.9

Hiring Teachers 3.9

Capacity for selectivity 3.10
School autonomy 3.12
Range of assessment methods used 3.13
Purposes and uses of assessments 3.16
Technical quality 3.17

Continued Employment: Retention and Tenure of Beginning Teachers 3.18

Range of methods used 3.20
Purposes and uses of assessments 3.23
Technical quality 3.25

Support 3.25

Range of methods used 3.26
Purposes and uses of assessments 3.28
Technical quality 3.29

Focus of Assessments 3.29

Subject-matter Competence 3.30
Pedagogy 3.30
Content Pedagogy 3.31
Other Constructs 3.31

Differential Impact of Assessment Practices 3.33

Costs of Assessments 3.33

Hiring Assessment Costs 3.34
Continued Employment Assessment Costs 3.35
Support Assessment Costs 3.36
Contributions of LEA Assessments to Staff Development Decision 3.36

Summary and Conclusions 3.37

Summary of Major Findings 3.37
Conclusion 3.40

CHAPTER 4: THE STATE ASSESSMENT PRACTICES SUBSTUDY 4.1

Description of Sample and Methods 4.1

Sample Description 4.1
Methodology 4.2

State Context 4.2

Basic Skills Proficiencies Requirement 4.3
Subject-matter Knowledge Requirement 4.3

Multiple subjects credential 4.4
Single subject credentials 4.5
Trends in the use of examinations to demonstrate subject-matter knowledge 4.6

Review of Credential Applications 4.6

Overview of the Major Assessment Point 4.8

Range of Methods Used 4.8
Purposes and Uses of Assessments 4.10
Technical Quality 4.10

Document review 4.10
Standardized tests 4.11

Focus of Assessments 4.12

Subject-matter Knowledge 4.12
General Pedagogy and Content Pedagogy 4.14
Basic Skills Proficiencies 4.16

Differential Impact of Assessment Practices 4.17

Ethnic groups 4.18
Gender groups 4.20
Type of credential 4.20
Review of studies of potential sources of differential performance 4.20

Costs 4.23

Contribution of State Assessments to IHE and LEA Decisions 4.23

IHE Decisions 4.23
LEA Decisions 4.25

Summary and Conclusions 4.25

Assessment Context 4.25
Methods of Assessment 4.25
Focus of Assessments 4.25
Technical Quality of Assessments 4.26
Assessment and Support 4.26
Conclusions 4.27

CHAPTER 5: EXPERIENCES OF BEGINNING TEACHERS 5.1

Sample Description 5.1

Teachers 5.1
Range of IHEs and LEAs Represented 5.4

Traversing the Credentialing Assessment System 5.4

Subject Matter Preparation in College 5.5
Admission to a Teacher Education Program 5.6
Education Courses Prior to and During Student Teaching 5.7
Admission to Student Teaching 5.8
Student Teaching 5.9

Assessment by university supervisors 5.11
Assessment by master teachers 5.13
Overall assessment experience 5.16

Employment Assessments 5.21
First-year Assessments 5.22

Assessment Content 5.30

General Pedagogy 5.30
Content Pedagogy 5.32
Working with Diverse Students 5.34
Curriculum Frameworks 5.36

Summary and Conclusions 5.38

The Forms of Assessment 5.38
Characteristics of Assessors 5.38
Content of Assessments 5.38
Rigor of Assessments 5.40
Thoroughness of Assessments 5.40
Redundancy of Assessments 5.41
Consistency of Assessments 5.41
Fairness of Assessments 5.42
Helpfulness of Assessments 5.42
Assessments and Support 5.43

CHAPTER 6: SUMMARY OF FINDINGS ACROSS SUBSTUDIES AND CONCLUSIONS 6.1

Focus of Assessments 6.1

Subject-matter Knowledge 6.1
General Pedagogy 6.2
Content Pedagogy 6.2
Knowledge of Students and Student Diversity 6.3
General Academic Ability and Personal Attributes 6.3

Forms of Assessments 6.3

Technical Quality of Assessments 6.4

Extent to which IHE, LEA, and State Assessments Constitute an Assessment System 6.5

Rigor 6.5
Thoroughness 6.7
Fairness 6.9
Consistency 6.9
Helpfulness 6.10

Conclusions 6.11

BIBLIOGRAPHY B.1

CHAPTER 1:

INTRODUCTION

CHAPTER 1:

INTRODUCTION

The need for better preparation of beginning teachers is both a major national and State policy issue. As part of broader discussions about professionalizing teachers, researchers, and policymakers alike have argued that more rigorous and comprehensive assessments of teachers' knowledge and competence should be developed and adopted similar to the standards of other professions. For example, the National Board for Professional Teaching Standards is proceeding with the development of assessments for the certification of accomplished master teachers. Simultaneously, other research has pointed to the need for stronger support systems for beginning teachers. One major cause often cited for high attrition rates among beginning teachers is the exhaustion and sense of inadequacy that accompanies the transition from student teaching to first year teaching. Typical work conditions isolate teachers from their peers, providing new teachers with limited opportunities to develop realistic standards for their performance. Overall, findings from varied strands of research and analyses point to the need for more comprehensive systems of support and assessment during teacher preparation, credentialing, and the first years of teaching that recognize the developmental nature of teachers' professional growth in order to promote the successful induction of teachers into an effective and respected profession.

Moreover, as the composition of the nation's public school population becomes more culturally, ethnically and linguistically diverse, the task of preparing teachers also to deal effectively with diverse student groups and needs is especially challenging. This need is particularly critical in California where student enrollment is growing by 225,000 pupils annually. With a school population of 3.5 million that embraces over 94 languages and dialects, one in six children in the system as a whole does not speak English. Adding to the challenge, this increasingly different and growing population is, and will be, spread unevenly across one of the largest and most geographically varied states.

While these demands suggest a need for different training for teachers, the general increase in population and other demands placed on the state human service sector have forced State policymakers to make difficult decisions about limited resources. Thus, California is faced with the daunting challenge of crafting standards for its teaching force that reflect the new demands on its public schools while managing to stay within the limits of available resources. This will call for "working smarter rather than harder." Moreover,

California must provide support for the teachers selected to work in its public schools so that the efforts made to prepare and recruit them are not wasted. Additionally, the assessment of new teachers must be fully responsive to these new and heightened demands while addressing the call for rigorous, thorough, fair, consistent, and helpful evaluations.

In response to legislative and gubernatorial mandates that began in 1985 with the California Commission on the Teaching Profession and culminated in the passage of SB 148, the Bergeson Act, the California New Teacher Project (CNTP) was created and charged with exploring innovative methods of new teacher support and assessment. The CNTP has three components: support, evaluation, and assessment. Briefly, the support component of the CNTP consists of local pilot projects that represent diverse teaching contexts as well as a variety of approaches to supporting new teachers. The evaluation component of the CNTP is designed to investigate the effects of the various methods of support on new teacher effectiveness and retention, as well as cost effectiveness. The assessment component in its first two years was designed to develop and pilot test innovative forms of new teacher assessment. Several other reports discuss the findings and analyses of these three components. This document reports the work in the third year of the assessment component of the CNTP.

Assessment Component of the CNTP: Phase III

In the past three years, the California Department of Education (CDE) and Commission on Teacher Credentialing (CTC), the two State agencies charged with developing new teacher policy, have conducted research on the support and assessment of new teachers. The research on support is described in reports disseminated by the Evaluation Component of the CNTP. The evaluation of diverse approaches to teacher assessment was intended to identify the most promising ways in which a comprehensive assessment of teacher candidates could inform the credentialing process and improve the quality of teaching in California public schools. In the first two years, research focussed on pilot testing innovative forms of new teacher assessment. In planning the research to be conducted in the third year of the assessment component, staff from the CTC and CDE considered an understanding of the *current* system of evaluating new teachers, including gaps, redundancies, and the technical quality of assessments currently in place to be critical to guide the development of any alternative models of assessment for new teacher policy. Indeed, the advisory panel to the CNTP argued that a careful evaluation of existing teacher assessment practices was crucial to understanding whether major reforms were needed and to determine the impact of particular reform efforts undertaken by teacher education programs and school districts in the past decade.

There is growing recognition that the knowledge base of teaching is very complex (Shulman, 1987) and that different skills are likely to develop at different stages of a teacher's career. The emerging literature on differences between new and experienced teachers suggests that completion of teacher preparation programs, no matter how well structured, cannot fully prepare teacher candidates to perform as excellent practitioners in the classrooms (Berliner, 1989), and that some skills may be present in only rudimentary form in new teachers.

To identify promising and realistic alternative assessment systems and policies, proposed revisions in assessment practices need to take into account:

- the developmental nature of teacher professional growth; and
- the domains of knowledge and skills that should be assessed in relation to the domains of knowledge and skills that existing assessment practices currently do assess.

However, knowledge of the current range of assessment practices for new teachers and the constructs they measure is fairly limited. Few research studies have been done on this topic, especially in the realm of the technical quality of assessments (e.g., training of assessors, reliability of the methods used). While CTC program reviews provide some information on the assessment practices employed in teacher preparation programs, the information is not current for all programs (which are evaluated once every six years) nor is it provided in sufficient detail either to evaluate how constructs are operationalized or to identify the standards that are applied. Even less is known about the range of district assessment practices for beginning teachers across the diverse teaching contexts represented within California's public schools. In addition, issues of whether and how well different assessments are linked and provide useful information need to be more fully examined. Finally, both preservice teacher education programs and school districts have initiated new efforts at support and assessment through internal reforms and external grants. No comprehensive evaluation of the cumulative impact of these changes has yet been made in California. A careful Statewide review of the state of assessment practices was seen as an important element in any proposed plans for further reform in new teacher support and assessment. Before highlighting the specific elements of this study of existing assessment practices, an explanation of the context in which teacher assessment takes place is important.

Current Context for Assessment of California Teachers

To be a teacher in California involves academic preparation in an appropriate subject, professional preparation, and successful review by the hiring school district. Specific requirements for a California preliminary teaching credential for the regular elementary and secondary classroom teacher are as follows:

- Passage of the California Basic Educational Skills Test (CBEST);
- Possession of a bachelor's degree;
- Demonstration of subject matter competence through one of the following:
  - Passage of a relevant subject matter examination, or
  - Completion of an approved an approved waiver program (a course of study in the subject or subjects to be taught), offered by a California college or university or an equivalent program offered outside California;
- Completion of a program of study in professional preparation meeting California standards;
- Obtaining a certificate of clearance affirming either an absence of criminal convictions or, in the case of some convictions, rehabilitation;
- Completion of a reading methods course or examination, and
- Completion of a course or examination covering the United States Constitution.

A preliminary credential is valid for five years and is nonrenewable after the initial five-year period. For a renewable credential, called the professional clear credential, a candidate must complete a fifth year of study, and complete instruction in health (including alcohol, drugs, tobacco, and the study of nutrition), mainstreaming issues in education (PL 94-142), and computer education. The fifth year of study (approximately 30 semester units or 45 quarter units) can be done in another credential program (special education, educational administration, etc.), a graduate degree program, or an individualized program of study approved by an IHE. To renew the professional clear credential, a teacher must

complete a professional development plan equalling 150 clock hours of activity every five years. The range of permissible activities is great, and the verification of the completion of these activities is the province of the school district.

About 75% of the candidates doing their teacher preparation in California complete the initial certification as the fifth year of higher education. It is possible to take the additional instruction beyond teacher preparation that is mandated by law for the professional clear credential concurrently with the teacher preparation coursework and receive a professional clear credential as the first credential. Otherwise, teachers have up to five years to complete this instruction.

Candidates who received their teacher preparation as part of their undergraduate degree (i.e., out of state) cannot qualify immediately for the professional clear credential. These candidates need to work on their fifth year of study as well as any of the mandated instruction not completed, within a five year period.

The Commission on Teacher Credentialing has approved basic teaching credential programs at 21 campuses of the California State University system, 8 campuses of the University of California system, and 44 private and independent colleges and universities. While the internal organization of these programs vary, all programs are required to meet standards of program quality established by the Commission. These standards, and other existing State laws, determine the required experiences and courses for all credential candidates. These current standards have been in effect since 1986 and reflect movement from a compliance orientation to one of broader, qualitative standards. All credential programs are reviewed by an outside team of educators drawn from the ranks of college faculty, credential holders, and other educators. These reviews include document analysis and site visits which include interviews with students, graduates, and employers. In addition to these reviews of professional preparation programs, the Commission also sets standards for subject-matter knowledge. Two principal means of demonstrating subject-matter competence exist. The first involves passing the authorized subject matter knowledge examination(s) at the score level set by the State. The second is to complete a program of study at an accredited institution that has had its program of study approved by the Commission. The review by the Commission is to ensure that the program of study covers the knowledge and skills contained in the State curriculum framework for that subject. The Commission has recently revised these standards to be more congruent with the needs of contemporary schools. Future approvals of such programs of study will involve site visits by teams of trained reviewers. Other than the examination of candidates using the California Basic Educational Skills Test (CBEST), which examines reading skills, basic mathematic

skills, and writing ability, and the examination of those candidates who must take the State adopted examination of subject-matter knowledge for that credential, California preservice teachers are not individually assessed by the State. Rather, the credential program they attended is assessed.

For the more than 1,000 California school districts, the assessment of teachers in service is direct and conducted under local standards and expectations. Hiring practices are also locally controlled as school districts vary in their needs and perceptions of who will "fit" with the district's mission and goals. The typical pattern is for districts to advertise their vacancies when they occur and to seek qualified candidates from local colleges and universities. The recent rise in school enrollments in California, coupled with the increasing need for teachers with bilingual and ESL skills, has forced some districts to seek candidates from outside the State and the nation. Additionally, hiring patterns are driven largely by student enrollments so that some districts hire one or two candidates a year while larger urban districts may have to hire 2,000 teachers in a year. Districts also vary on who handles the hiring process. In large districts, there is often a district personnel office that screens candidates and sends qualified candidates out for interviews. In smaller districts, the superintendent may do all the hiring or it may be delegated to a site principal. A number of districts in California must also balance racial integration issues with instructional needs. Finally, recent economic difficulties have made it more difficult for school districts to increase wages which would make teaching more financially rewarding. All these contextual issues have made it more difficult to obtain a clear sense of the level of assessments, their technical quality, relation to prior assessments, and connection to the actual work of teachers.

In reviewing the types of assessments teachers experience in credential programs and in school districts, it became clear that some assessments were simple requirements that had little evaluative information about the candidate (e.g., all candidates must pass a course in the U.S. Constitution). Other assessments were complex and potentially evaluative about the skills and knowledge of the candidate, such as formal observations during student teaching, or evaluations of performance during the employment interview. It was evident that those assessments which focussed on teacher performance with students received greater weight and concern than other assessments which were oriented to issues outside the classroom (e.g., requirements such as the required health course). The variations in level of assessments and importance attached to the outcome of such assessments were used to shape the data collection design and, in particular, the questions asked in the interview protocols.

Essential Design Features of this Study

This study was designed to yield information about existing assessment practices at college and university level, the local school district level, and the State level within the context of the legal, social, and economic situation facing California educational systems. Through the analysis of such existing assessment practices, the domains of knowledge and skill that are considered important by the various constituent groups (e.g., teacher educators in colleges and universities, local school district administrators, school board members, and others involved in teacher assessment activities) would be revealed whether they were part of the formal assessment practices or part of the attendant values that undergird assessment in education. This study was also planned to provide information that would lead to the conceptual base and design of a new system of support and assessment for new teachers in California. Finally, the findings of this study will also assist in the development of two other elements in the emerging policy framework: 1) an analysis and development of a conceptual framework that defines the important components of good teaching, and 2) an analysis of alternative assessment systems and policy options for the improvement of teacher credentialing decisions. This volume is the final report and analysis of the set of studies designed to describe the range of existing new teacher assessment practices.

In reviewing the research literature on teacher assessment and the policy concerns raised in the reform literature, several principles of teacher assessment emerged that would signify an ideal system (Figure 1.1).

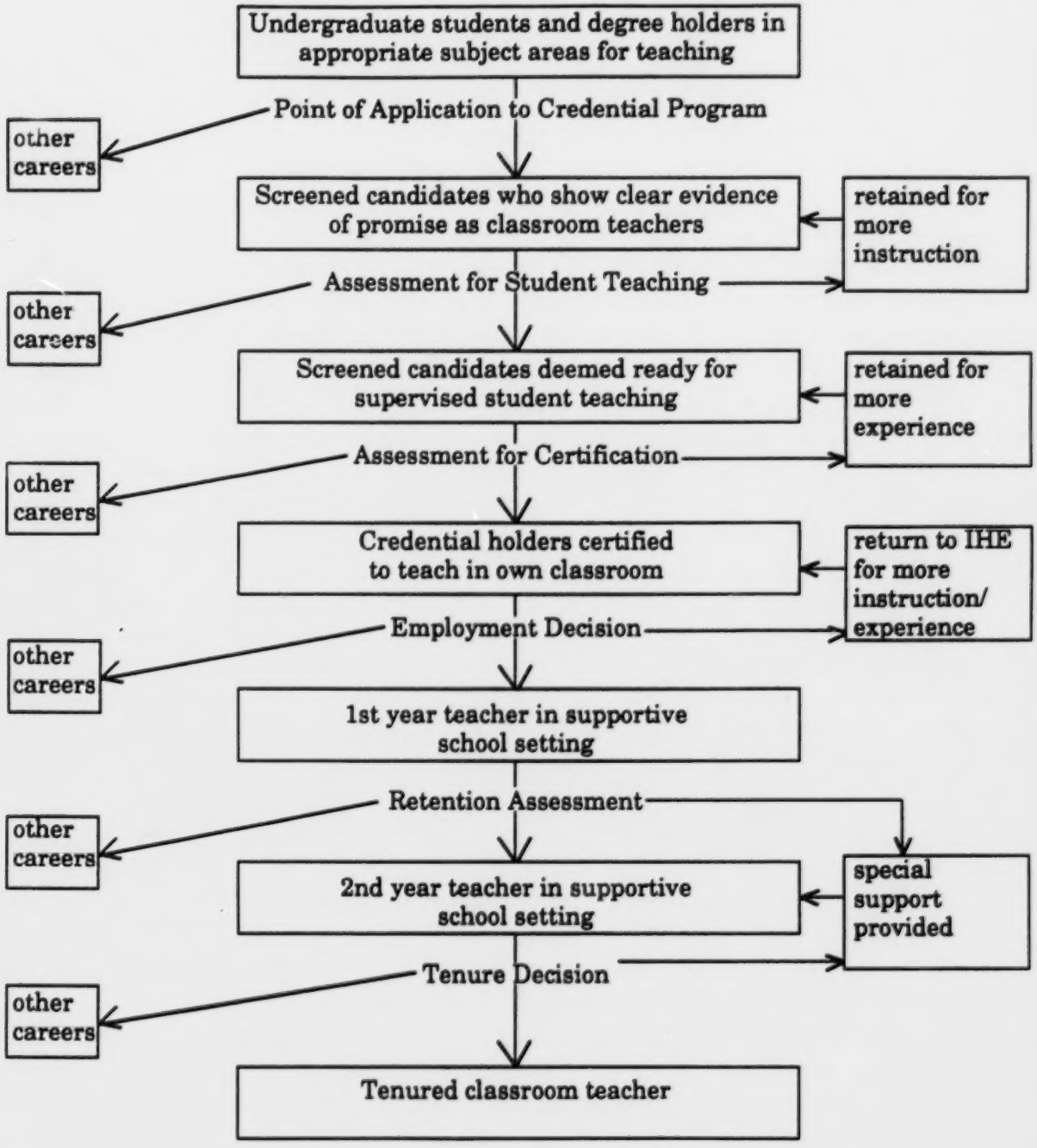
Guiding Constructs for an Ideal Statewide Assessment System

The following are constructs which we propose as characteristics of an ideal assessment system. These constructs guided the design and analysis of the study and the construction of the interview protocols.

- Rigor

Assessments should include high but reasonable levels of expectations, clearly defined criteria and standards, and procedures for assuring reliability of assessments. At a minimum, the procedures for assuring reliability should include training and monitoring of assessors.

FIGURE 1.1
IDEALIZED TEACHER ASSESSMENT SYSTEM


Undergraduate students and degree holders in appropriate subject areas for teaching
Point of Application to Credential Program
other careers
Screened candidates who show clear evidence of promise as classroom teachers
retained for more instruction
Assessment for Student Teaching
other careers
Screened candidates deemed ready for supervised student teaching
retained for more experience
Assessment for Certification
other careers
Credential holders certified to teach in own classroom
return to IHE for more instruction/ experience
Employment Decision
other careers
1st year teacher in supportive school setting
Retention Assessment
other careers
2nd year teacher in supportive school setting
special support provided
Tenure Decision
other careers
Tenured classroom teacher

- Thoroughness

Assessments should be guided by a conceptual framework which includes all major aspects of teaching. Diagnostic assessments should utilize information from previous assessments to measure progress. Assessments should not be redundant; where the same skill is focussed on at a later career stage, growth should be expected.

- Fairness

Assessments should take account of the anticipated stage of development, recognizing that competencies develop at different rates, with some competencies requiring experience with a variety of children or topics to develop a basic level of competence. Assessments should also apply to teachers in different contexts.

- Consistency

Assessments Statewide should use common criteria, although local circumstances may dictate additional criteria. The minimal frequency of assessment, timing of assessments (e.g., beginning, middle, and end of the year), and extent of assessments (e.g., length of an observation) should be similar for all prospective and beginning teachers making reasonable progress. (Teachers who are having difficulty should expect more frequent assessment.) The shifts in the focus of assessment (e.g., the specific level of competence in each skill or knowledge competency) should be similar for teachers making major career transitions, e.g., entering student teaching or the first year in teaching.

- Helpfulness

To contribute to teacher development, assessments should provide for both formative and summative feedback, and identify both strengths and weaknesses, so that prospective and beginning teachers can take pride in their accomplishments and work on their weaknesses. For each desired skill, the assessments should describe where a teacher is along a developmental continuum. The assessments should also be coordinated with some form of support for developing skills.

Development of Interview Protocols

In addition to the guiding constructs, the development of the interview protocol created for collecting data in this study began with developing questions on the following dimensions of existing assessment practices:

- The aspects of teacher expertise, competence, and performance that are currently evaluated in new teacher assessment practices, the ability of these systems to measure the knowledge and skills considered of greatest importance, and the standards used to measure those criteria.

- The impact of existing assessment practices on different candidates, particularly candidates of diverse ethnic, linguistic, and academic backgrounds.

- The diversity of existing practices in terms of representing the varied and distinct missions of institutions of higher education (IHEs) where most new teachers are prepared, and the distinct constituencies of different local education agencies (LEA) where most new teachers begin their careers.

- The overall coherence and completeness of the assessment practices in combination with each other.

- The extent to which existing assessment practices produce information that is useful for prospective and beginning teachers.

In examining current assessment practices, prospective and beginning teachers are assessed at eight formal points during their career preparation and development:

- Assessment of subject-matter knowledge in undergraduate coursework

- Assessment of applicants for admission to professional teacher education

- Assessment of candidates prior to advancement to student teaching

- Assessment of performance in student teaching and teacher education courses

- Assessment via CBEST and other State-mandated examinations to satisfy credentialing requirements

- Assessment of applicants for employment as certificated teachers
- Formative assessment of new teachers by the LEA for support, and
- Summative assessment of new teachers by LEA supervisors for tenure.

Table 1.1 shows the connections among the ordering of these formal assessments points and the individuals and organizations currently performing the assessments. Since the three organizations, i.e., colleges and universities (IHEs), local school districts (LEAs), and the State licensing agency (CTC), all play a crucial role in the total assessment experience of a teacher yet do not formally share any single assessment point, four substudies were conducted to obtain the depth and breadth of understanding necessary to make recommendations about teacher assessment reform.

In order to describe and evaluate the range of assessment practices currently in use in California, the first three substudies focussed on the practices of IHEs, LEAs, and the CTC. The overall coherence and comprehensiveness of these assessment practices as well as their ability to produce information that is useful for prospective and beginning teachers was the focus of the fourth substudy that examined the assessment activities as they were experienced by a sample of individual beginning teachers.

Given the diversity of teaching contexts in California and the distinct missions of the various IHEs that prepare teachers, the IHE and LEA studies needed to be designed to capture natural variation in local settings and conditions, while simultaneously maximizing the diversity of assessment practices. At the same time, characteristics of the State and local context help explain why particular IHEs and LEAs choose specific assessments and set the parameters for any new alternative assessment systems. Specifically, the context, larger purposes of the project, and research issues posed for this task suggested the following points in the research design for each substudy:

1. Though each substudy's primary focus should be on the description and analysis of prospective and beginning teacher assessments currently in use, the factors that influence current, former, and future practices also need to be addressed to understand why current assessment practices are the way they are.

TABLE 1.1

PROVIDERS OF ASSESSMENT

Assessment Points	Teacher Education Faculty	Department Faculty	University Supervisor	District Personnel	School Personnel	Cooperating Teachers	State of California
Undergraduate Coursework		X					
Admission to Credential Program	X	X					
NTE/CBEST							X
Admission to Student Teaching	X	For some single subject programs	X			X	
Credential Recommendation	X	For some single subject programs	X			X	
Employment as 1st Year Teacher				X	In some districts		
Retention after Year 1					X		
Tenure Decision				X	X		

2. Given the exploratory nature of several aspects of the study, the substudies should not be designed to make comparisons between participating sites or select a statistically representative sample. Instead, each substudy needed to sample IHEs and LEAs with characteristics that span across many dimensions relevant to policy to analyze the conditions under which certain assessment practices are likely and certain assessment constructs more important in order to make reasoned judgements about the generalizability of specific findings.

3. This type of analysis required a purposive sampling plan to select sites that cover the range of natural variation in conditions and settings, while maximizing the diversity of assessment practices.

4. Both the limited knowledge base about which domains of knowledge and skills are important, which are actually assessed, and the potential sensitivity of assessment practices to local contexts imply the need for field investigations to ensure that the scope and subtleties of the information needed could be obtained.

The basic data collection design called for the selection of a sample of 12 IHEs and 12 LEAs with on-site interviews with administrators, faculty/teachers and other staff, and the collection of assessment documents and materials from each site for a complete list of documents requested). In general, sites were selected to capture the range of local settings as well as a diversity of assessment practices. Selection of sites and persons to interview as well as the specific data collection procedures are discussed in the respective chapters on LEA and IHE assessment practices.

The interview protocols were developed using the guiding principles of an ideal assessment system, information from teacher education experts and school district researchers explaining who to interview about the eight assessment points, and the four research design considerations. The final array of interview protocols was based principally on the typical workload of key individuals in the colleges and universities and the local school districts. Each protocol reflected the guiding principles but was focussed on the particular assessment activities of that individual. For example, the longest interviews for the IHE substudy were with program directors who are faculty members in charge of credential programs. Since they have responsibilities for every part of the credential program, they could give useful information about each assessment point in the credential program. Conversely, the interview protocols for evaluators of subject-matter competence

focussed on the determination of candidate knowledge of the subject(s) to be taught. Similar variations existed within the LEA substudy.

In order to assess the overall coherence and comprehensiveness of these assessment practices as well as their ability to produce information that is useful for prospective and beginning teachers, the fourth substudy was designed to examine these assessment practices as they were experienced by a sample of individual beginning teachers. To ensure that teachers with a range of knowledge and skills were included and to compare the findings of the existing assessment practices with the results of assessments pilot tested in phases I and II of the assessment component of the CNTP, the design called for a sample of 25 teachers with the following characteristics:

- All teachers had earned one or more of the seven basic California teaching credentials (Multiple Subject, Multiple Subject Bilingual Emphasis, Single Subject: English, Mathematics, Life Science, Physical Science, Social Studies).

- All teachers had been initially employed as beginning teachers by California school districts at some point between 1987 and 1990.

- All teachers had participated in the pilot studies of alternative performance assessments during phases I and II of the Assessment Component of the CNTP.

- All teachers provided consent so that institutional records of assessments of their work could be obtained from IHEs and LEAs.

The basic design called for in-depth interviews with each teacher and the collection of assessment records about each teacher from IHEs, LEAs and the CTC in order to examine teachers' perceptions of their assessments as well as to allow for independent analysis of the assessments with a focus on the overall rigor, thoroughness, fairness, consistency, and helpfulness of the existing assessments.

In the chapters that follow, each of these five substudies are described in relation to the guiding principles of an ideal assessment system for California classroom teachers. The report concludes with a summary of the overall coherence and completeness of the assessment practices currently in use in preservice teacher preparation and in school district hiring and retention policies from the perspective of a sample of the individuals who conduct those assessments and a sample of the teachers who have most recently experienced

them. This combination of perspectives will provide policymakers with a much clearer picture of the state of teacher assessment in California than has been heretofore available and should enable future policy changes in teacher assessment and support to take best advantage of this enhanced picture.

CHAPTER 2:

INSTITUTIONS OF HIGHER EDUCATION ASSESSMENT PRACTICES

CHAPTER 2:

INSTITUTIONS OF HIGHER EDUCATION ASSESSMENT PRACTICES

This chapter discusses the major findings from the study of how institutions of higher education (IHEs) with approved teacher education programs evaluate candidates for admission to the professional preparation program, assess candidates for retention during the program of study, with particular attention paid to advancing candidates to formal student teaching, and determine which candidates will receive institutional recommendation for a basic teaching credential in California. Since the State of California makes credentialing decisions based on the recommendations of faculty in approved programs of study, understanding the practices and policies of these pre-service programs will help in determining the level of knowledge and skills expected of teachers upon departure from their professional preparation.

The chapter is divided into eight sections. The first describes the sample of institutions selected for the study and the methodologies employed to gather and analyze the data. The second portion of the chapter presents a brief overview of the context in which teacher preparation operates in California. The third section reports on the major assessment points as candidates enter and complete the teacher preparation programs. The fourth section discusses the data from the perspective of the focus of the various assessment activities while the fifth section reports on the differential impacts of each assessment point on the individual candidates. The sixth section addresses the costs to institutions of each of the major assessment points in the basic credential programs. The seventh portion of the chapter reviews the contribution each assessment practice makes in determining who is ultimately recommended for a credential. The final section presents a summary of the findings and some conclusions about the status of pre-service assessment activities currently employed by California teacher education programs.

Description of Sample and Methods

Sample Description

Twelve institutions of higher education with approved basic credential programs were selected to represent the range and distribution of teacher preparing institutions and credential programs in California. These institutions account for only 16% of the total

number of colleges and universities preparing teachers in the State, but the sample was purposively constructed to include those types of institutions likely to reflect both innovative practices and typical patterns of assessment. Thus, 7 of the institutions were drawn from the 21 campuses of the California State University system as it is the largest preparer of teachers in California (70% of the total number of institutionally recommended candidates). Three institutions represented the private and independent colleges and universities of California. Although 45 such institutions actually prepare teachers, their programs tend to be small so that the private sector only accounts for 20% of the institutionally recommended candidates for credentials. Finally, 2 institutions from the University of California system were included. This branch of public higher education accounts for 10% of the teachers prepared in State.

In addition to selecting the 12 institutions on the basis of proportional representation, care was taken to select campuses that reflect the extraordinary range of populations served in California and the wide array of communities that teachers must be prepared to serve. Thus, not only was attention paid to identifying urban and inner city campuses, but also rural campuses were included in the sample.

TABLE 2.1

TYPE OF INSTITUTION

	CSU Campus	P&I Campus	UC Campus
Inner City	02	01	01
Urban	03	02	00
Suburban	00	00	01
Rural	02	00	00

Beyond issues of demographic concern, the sample of institutions was intended to capture some sense of the range of assessment practices current teachers would experience. Thus, information derived from Commission on Teacher Credentialing program evaluation reports was used to select those institutions that would fit both the demographic pattern desired but would give evidence of effective and "cutting edge" assessment of entering candidates to a teacher education program. Of particular interest were assessment practices at the time of admission to teacher education programs, assessment instruments and

practices for retaining only those candidates likely to be successful in student teaching, and, finally, those assessment instruments and practices used to make the final recommendation for certification as a beginning teacher.

Methodology

While teacher education programs in California are regulated by the Commission on Teacher Credentialing and, therefore, share many procedural similarities, the range of assessment practices across the approved basic credential programs dictated several means of collecting data. In addition to collecting printed information about the programs, site visits were proposed so that detailed interviews with key administrators and faculty could be conducted. These interviews were designed to provide information about not only the official polices and procedures related to assessment, but also the informal practices that surround these programs and indicate the values and philosophy of the programs.

Once the list of institu[illegible]s was approved by the study sponsors, each selected institution was invited to pa[illegible]cipate in the study through a letter sent by the Commission on Teacher Credentialing. Two of the original institutions declined to participate and back-up institutions were contacted. All participants were assured of anonymity through aggregate reporting of data. Moreover, there was no interest in comparing institutions. The intent of the study was to show the range of assessments currently in use in California teacher education programs. Since the protocols were not approved for use until the summer vacation had begun, more of the interviews took place over the telephone than had originally been envisioned. While this necessitated some modification of the interview protocols, none of the interviewers reported any serious problems with data collection and, in the case of the cooperating (or as referred to here, master) teachers, no other method would have worked during this time of the year.

Interview protocols. The interview protocols were created by reviewing State regulations on teacher education programs and by discussing with teacher education experts the range of organizational patterns in California credential programs. Those conversations resulted in a list of role types who would have knowledge of the assessment practices of basic credential programs and, in turn, helped to generate appropriate questions for the protocols.

Role types identified for interviews:

- Credential Program Directors/Coordinators: Multiple Subject, Single Subject, and Bilingual Emphasis
- Admissions Office Personnel
- Faculty Evaluators of Subject Matter Competence: Liberal Studies, Mathematics, English, Science, Social Science
- University Supervisors of Student Teaching (in above areas)
- Master Teachers (familiar with particular program)
- Faculty with Innovative Assessment Practices
- Faculty/Staff in Post-Graduate Follow-up

Because admission to, retention in, and departure from credential programs turns on a number of issues and because teacher education programs frequently involve faculty from outside individual credential programs and schools of education, five interview protocols were developed to capture the wide array of information necessary to address the research questions posed by this study.

1. **Administrative Procedures and Selection Criteria.** Program Directors or Coordinators (those faculty members charged with the administration of the approved credential programs) received the longest (typically three hours) and most detailed interview protocol as they usually possessed the most information about program policies and practices. It covered all major assessment points (admissions, advancement to student teaching, and recommendation for a credential), credential program policies (e.g., special admissions and reconsideration of admission, removal from the program), and programmatic issues related to staffing of programs and costs of conducting these assessments. If the institution was structured such that there was a central admissions office for all teacher education programs, the director of that office was also interviewed. This interview focussed on the handling of admissions across various programs and identified those practices that were common to all three basic credential programs and those practices particular to each. The Program

Director/Coordinator was asked all other questions related to administrative procedures for each program.

2. **Evaluations of Subject-Matter Competence.** Those faculty members charged with assessing the subject matter knowledge of each candidate were interviewed about the methods used to determine the candidate's grasp of the subject or subjects to be taught. These faculty are sometimes involved in the teacher education program directly or may be regular faculty within a department and be assigned by that department to conduct such assessments.

3. **Evaluations of Student Teaching.** To understand the ways candidates are assessed during student teaching, both university supervisors, college employees assigned to observe and evaluate student teachers, and master teachers, classroom teachers in local school districts whose classrooms are used for the practice teaching experience and who provide much support and training to the student teacher, were interviewed regarding their perceived roles and expectations in the preparation program. As with the subject matter evaluators, attention was paid to interviewing individuals from each of the selected subject matter areas.

4. **Innovative Performance Based Evaluations.** If preliminary interviews with senior administrators revealed unusual types of assessment during teacher preparation course work, but distinct from student teaching assessment and subject matter knowledge assessment, a special interview was set up with the faculty member(s) responsible for those assessments. This interview focussed on the development of alternative assessments of subject matter knowledge and pedagogical knowledge other than paper and pencil examinations.

5. **Post-graduate Follow-up.** Finally, efforts were made to determine the kinds of post-graduate follow-up practices conducted by teacher preparation programs by asking about their support programs for graduates and the patterns of strengths and weaknesses noted in their graduates.

Documents collected. Lastly, each institution was asked to provide copies of all relevant materials that would explain the processes and procedures employed in their multiple subject, single subject, and bilingual teacher education programs. (See Figure 2.1.)

FIGURE 2.1

IHE DOCUMENT CHECKLIST

IHE: _____

Program: ___ Multiple Subject ___ Bilingual/Crosscultural

___ Single Subject ___ Admissions

I. Admission

___ Admissions packet

___ Multiple Subject Program

___ Single Subject Program

___ Bilingual/Crosscultural Program

___ Official policies and procedures governing evaluation of the admission packets

___ Forms for evaluation of the admission packets

___ Document(s) defining admissions requirements

___ Procedures for making admissions decisions

___ Policy governing handling of incomplete packets

___ Procedures or guidelines to orient people who evaluate the admissions packets

___ Policy and procedures for conducting and evaluating interviews with applicants

___ Standardized interview protocol for applicants

___ Form for summarizing and/or evaluating interview

___ Procedures or guidelines to ensure reliability among people conducting interviews

___ Policy and procedures for synthesizing all information obtained about applicants into an admissions decision

___ Policy regarding weighting information about a candidate

___ Form for summarizing a candidate's qualifications for admission

II. Evaluation of Subject-Matter Competence

___ Form(s) for evaluating subject matter competence of applicants for the following subjects:

_____

_____

___ Waiver program requirements for the same subjects

III. Evaluations Prior to Student Teaching

___ Policy and procedures for admission to student teaching phase of program

___ Application form for admission to student teaching

IV. Student Teaching

___ Policy and procedures for selecting cooperating teachers

___ Policy and procedures for selecting university supervisors

___ Policy or procedure for formative evaluation of student teachers

___ Form for formative evaluation of student teachers

___ Document describing training of university supervisors and/or cooperating teachers

___ Policy or procedure for evaluations of student teachers using _____ (non-observational method)

___ Form for evaluating _____ produced by student teachers

___ Policy and procedures governing summative evaluation of student teachers

___ Form for summative evaluation of student teachers

___ Policy or procedures for handling marginal student teachers or for requiring additional units of student teaching

___ Policy or procedures for resolving differences between the university supervisor and the cooperating teachers in the evaluation of specific student teachers

V. Recommendation for a Credential

___ Policy and procedures for recommending candidates for a credential

___ Form for recommending a candidate for a credential

These materials ranged from informational flyers and brochures about the institution and its programs to highly detailed handbooks prepared for students and master teachers regarding the expectations and requirements for student teaching. Of particular interest to the interview teams were those evaluation instruments used in assessing candidates for admission, retention, and, especially, recommendation for the initial credential.

Interview teams. Five teams of trained interviewers (three teams visited more than one institution) made campus visits to conduct these interviews. (See Tables 2.2 and 2.3.) Teams were composed of two or three members, depending on institution size. Team members were experienced educators or graduate students with backgrounds in qualitative research. Each was trained in the use of the interview protocols by spending a full day of training reviewing the instruments and participating in simulated interview activities.

TABLE 2.2

INTERVIEWS CONDUCTED BY INSTITUTION AND TYPE

	CSU Campus	P&I Campus	UC Campus	TOTAL
Prog. Dirs	19	06	05	30
Other Adm	05	03	00	08
Subj Evals	19	06	05	30
Uni Sups	21	11	07	39
Coop Teach	30	08	06	44
Av # of Docs	24	15	24	21

TABLE 2.3

INTERVIEWS WITH FIELD SUPERVISORS BY SUBJECT AREA

	M/S Waiver	English	Math	Sciences	Soc Sci	Biling
Uni Sup	20	06	06	04	03	00
Coop Teach	22	08	07	04	03	01
TOTAL	42	14	13	08	06	01

The interview teams synthesized the data received from the various interviews conducted and documents submitted by the institutions into a site summary review form for each institution. This form requested the team members to make both qualitative judgments on the context of the programs reviewed and quantitative statements regarding such aspects of admissions and assessment as faculty and staff time commitments. The full day of debriefing the interview teams following the site visits focussed on the generation of common themes and patterns of assessment practices. Since many programs use different terms for similar elements in credential programs and many institutions calculate faculty/staff time in different formulas, part of the debriefing exercises were devoted to finding common understandings and identifying areas of uncertainty for follow-up telephone calls. All of the site summary review forms were analyzed and further synthesized and the principal assessment instruments and practices were compared across the 12 institutions in the sample.

Institution of Higher Education Context

California institutions of higher education (colleges and universities) must obtain approval from the Commission on Teacher Credentialing, an independent State agency charged with all educator licensure matters, to offer programs of study leading to a State teaching credential. State law requires that all teachers in publicly supported schools (K-12) hold the appropriate credential for the work they do. Currently, California operates under an approved program model to ensure that these credential holders can perform their assigned tasks. The competence of the credential holder is determined by ensuring that the program which prepared the candidate meets or exceeds defined standards. Other than the CBEST (California Basic Educational Skills Test) and, for some candidates, the NTE

(National Teacher Examination), the assessment of candidate competence takes place indirectly.

The current standards for elementary and secondary classroom credential programs were implemented in 1986 after a period of extensive study and review by the Commission which began as a response to the national and State level concerns over teacher competence. The 32 standards are grouped into 5 categories: Institutional Resources and Coordination (6 standards), Admission and Student Services (5 standards), Curriculum (4 standards), Field Experiences (6 standards), and Candidate Competence and Performance (11 standards). Although Category 5 makes reference to candidate competence, the Commission has indicated that it is impossible to ascertain if every candidate possesses such abilities and knowledge but evaluation teams must make an effort to determine how the program faculty make that determination and if, in the judgment of the team, the program organization, curriculum, and field experiences will likely provide such competence. In addition to these standards, the Legislature has mandated certain requirements for all credential programs.

While students are permitted to complete teacher preparation along with the baccalaureate degree, over 75% of the candidates prepared in California do so after earning the baccalaureate degree. This tends to compress professional preparation into a one or two year program of study and reduces the integration of pedagogical study and subject matter study. Additionally, the recent increase in school enrollments, particularly in urban areas, has increased the number of teachers serving on emergency permits. These teachers are required to work on their professional preparation while teaching and present, therefore, a particular challenge to teacher education assessment practices.

In California, the student population is among the most diverse in the country. Cultural, racial, ethnic, and language diversity characterizes increasing numbers of schools in California. The demands made on beginning teachers by this level of diversity has, in the past, not received sufficient attention in pre-service teacher education programs. While credential programs are increasing attention to diversity as a teaching issue, the expectations for what beginning teachers can accomplish in diverse classrooms are not uniform within the profession.

Overview of Major Assessment Points

Admission to Credential Programs

In this section, the processes and criteria used to evaluate candidates at each of the major assessment points -- admission to credential programs, retention in credential programs, student teaching, recommendation for a credential, and post credential support -- are described. For each of these assessment points, discussion focuses on: (1) the range of assessment methods used, (2) the purposes and uses of assessments, and (3) the technical quality of assessments.

Range of methods used. Several methods of assessment are used during the admissions process which are discussed below.

Grade point average. Each credential program on each campus in the study sample (12 IHEs, 33 credential programs) collected information on how well the applicant did academically in prior college level course work. Typically, applicants submit transcripts showing all college work which are reviewed by the admitting individual(s). Since 1986, all institutions with approved teacher education programs are required to demonstrate that the average GPA of admitted students exceeds the average GPA of some comparable group of students. All CSU campuses have been required since 1985 to limit admission to teacher education to those candidates whose overall undergraduate GPA exceeds the median GPA for students in the same major at the admitting campus. These averages are calculated by the Chancellor's Office and re-calculated every three years to account for changes in grading practices. In those cases where an applicant attended another college or university, the system-wide GPA for that major can be used or the campus may impose higher GPA requirements. The exception to this rule is limited to 15% of the total pool of admitted students. This option was designed to account for academic "late bloomers," students who had difficulty making the transition from high school to college, and students who are not native English speakers but have skills needed for contemporary public schools.

Alternative assessments to GPA. As a part of the review of undergraduate preparation, questions were asked of university faculty about the kind of assessments done within the undergraduate degree programs. Obviously, faculty could only answer for their own institutions so the data are clearly incomplete as many candidates for credentials do not attend the same institution for baccalaureate degrees and teaching certification. Additionally, the subject evaluators were most knowledgeable about courses in their own

department and less certain about types of assessment conducted in general education courses across the campus.

All the sample institutions require all courses to conduct a formal final examination. All the respondents indicated that all courses required some type of mid-term examination. The most common type of examination was a combined test using essay questions plus questions using multiple-choice, matching, or short answer. The second most common examination was essay only. Less than 10% of the courses use portfolio or demonstrations as assessment devices and these are all laboratory courses and/or performance courses. Most undergraduate courses, according to those faculty interviewed, require students to do some type of writing and 10 campuses now have a formal upper division writing requirement or examination that must be completed. One institution has initiated a senior seminar project and the evaluators for the multiple subject credential at three different institutions specifically indicated that the new waiver programs for individuals seeking multiple subject credentials, implemented in Fall, 1991, will require a "capstone" seminar that would seek to integrate student learning.

Subject matter knowledge assessment. All programs in the sample gathered information about subject specific knowledge as well as general academic performance. In California, all teachers must demonstrate competence in the subject they will teach and do so in one of two ways. Passing the National Teachers Examination with a State mandated passing score confers competence. The alternate method requires a student to know, as a first or second year college student, that they want to teach. This second method sets out a series of courses in the teaching area that are congruent with the State curriculum framework and represents 45 semester units of course work (or more) which may be similar to institutional requirements for the degree but often involve specific courses required only of those who intend to teach in California. About a third of the sample institutions require students who present a passing score on the NTE as evidence of competence in the subject area to take 3-6 units of additional work to verify clearly the competence implied in the passing score. Along with the pre-professional field experience course now asked of credential candidates in 80% of the programs in the sample, those candidates needing to complete both the additional academic courses and the pre-professional field experience course can have their admission to a teacher preparation program delayed by a semester or a year.

Alternative assessment practices. In the California State University system, the determination of subject matter competence, whether by examination or by completion of an approved program of study, is to be verified by the appropriate academic department on

campus. This requirement, initiated in 1987, was designed to increase university responsibility for the preparation of classroom teachers by involving schools and departments other than education and to provide the necessary expertise for making complex decisions about competence in broad subject matter areas. For example, one English department in the study sample has a required course that all applicants for the single subject credential in English take prior to admission to the credential program. The candidates undergo a series of assessments designed to determine their ability to apply content knowledge in settings typical for classroom teachers. The candidates evaluate student writing and make recommendations for improvement. They also explicate poems and short stories and provide ideas for teaching such material to designated groups of students. They prepare and deliver a presentation on a topic related to the teaching of English. Finally, they engage in the writing process themselves and reflect upon their experience as learners. Candidates are evaluated by the instructors through a Likert scale for each exercise and must pass each exercise in order to receive a full recommendation for competence in their subject-matter.

All of the program directors and faculty who conduct the subject matter competence reviews in the California State University programs in the study sample (7 sites) referred to system-wide committees that had developed alternative assessment plans for each of the basic credential subject areas (Liberal Studies, Art, English, Foreign Language, Life Science, Mathematics, Music, Physical Education, Social Science). These alternative assessments, when fully implemented, will be part of the over-all admissions process. All of the proposed assessment plans involve some form of interview, some activities designed to determine the ability of applicants to talk about their subject matter in "teacherly" ways, and some type of comprehensive assessment of applicants' grasp of subject matter through "capstone" courses. The proposals also call for students to submit portfolios of prior course work that show the kind of subject matter analysis and synthesis of which they are capable. While the CSU campuses have the most elaborate subject matter assessment plans in the sample, the program directors and faculty who evaluate subject matter competence at the other five institutions stated that they also address subject specific knowledge, generally through the application interview.

CBEST assessment practices. By State law, all credential programs must require applicants to take the CBEST as a part of the admissions process and 60% of the sample programs in this study require a passing score on the CBEST before admission to teacher education. Two of the institutions that require a passing score on the CBEST do so because CTC regulations require passage before student teaching can begin and they feature early field work. All others require passage of the CBEST as a screening device. The remaining

campuses require passage of the CBEST before student teaching can begin but after professional education course work has begun.

Document assessment practices. All programs require letters of recommendation and some type of personal statement that accompanies the application. The purpose of the letters varies across the programs in the sample, but most are written by academic sources (faculty) or professional sources (employers) and focus on the general qualities of the applicant in both academic and work arenas. A few programs require the candidate to have the writer address specific topics or complete a Likert scale personality checklist, but the modal preference is for a traditional letter of recommendation. A few programs expect the letter of recommendation portion of the application to address the applicant's prior experience with children or youth while the majority of the sample ask for additional information in that category.

The personal statement also varies in purpose and direction. Some institutions are very directive and require a minimum length, stipulate that it be handwritten, and provide topics which must be addressed. Other institutions appear to accept more general statements about the applicant's desire to teach. A few of the application forms ask the candidate to reflect on their attitudes toward children and learning as well as to describe any relevant experiences with teaching.

Interview assessment practices. Interviews are required in all of the sample institutions and 33% of the programs require two interviews (typically one is with the academic department and one with the teacher education faculty). These interviews are conducted in a variety of ways but most feature one or more faculty members interviewing an applicant for 15 to 30 minutes using a set of general questions related to personal interest in and fitness for teaching. Any questions asked about teaching focus on generic pedagogical issues like classroom management. Two campuses reported including classroom teachers in the panel of interviewers. A third of the programs ask specific questions about experiences with different ethnic or racial groups or attitudes about cultural and linguistic differences. One campus in the sample made significant use of psychological tests to identify candidates with strong preferences for cooperative work patterns and high tolerance for ambiguous situations. This campus had trained observers perform these tests and rank-ordered the applicants on the basis of their performance. They also developed a telephone version of one of the tests to permit out of State candidates to provide the same information.

Knowledge of schools and students assessment practices. All institutions in the sample (and all credential programs in California) require applicants to submit evidence of

some type of experience with children/youth and 80% of the study sample have specified the number of hours of documented contact with children or youth (45 is the modal number with the highest being 100 hours) and specified that the time must be in a classroom setting, preferably a public school classroom. All the programs in the sample offer a preliminary field experience course to meet this requirement and some larger institutions have such courses embedded in the academic department so that their majors can experience field work in that particular subject area. These courses also require students to engage in work related observations like school board meetings, directed interviews with teachers and administrators, special education classrooms and the like. Post-graduate applicants can apply experience obtained elsewhere through letters of verification. A few institutions are now requiring more verification than just time spent and have the verifier fill out a form indicating what things were actually done or observed.

Summary admission review practices. Two-thirds of the programs did engage in some type of summary review after all the pieces of the application file were assembled. Those institutions that accept a specific number of students (33%) often attempt to rank order all applicants and use a scoring system which weights some elements over others. In those programs where all qualified applicants are admitted, the review is to ensure that the applicant has met the published criteria. Since CBEST, GPA, departmental approval of subject matter competence, and prior experience with children and schools are the most clearly defined of the published criteria, these appear to be the key variables for this sub-group of the sample. A few candidates may injure their chances of admission by poor performance on the general interview, but the more typical case is a student failing the CBEST, not meeting the minimum GPA or not meeting the subject matter approval requirements. For these students, the admissions answer from the program director is not a clear denial, but a decision of not ready for admission at this time. Thus, those students deemed not qualified by reason of missing courses are told to come back and re-apply when the additional courses have been completed. Some anecdotal information from program directors suggests that a call for additional experience with children or youth may be based on the perception of an incompatible personality with teaching. The notion is that, after experiencing a classroom first hand, the incompatibility will be apparent to the candidate who will not pursue teaching as a career and thus forestall the formal denial of admission.

Those programs which take a specific number of applicants and have more applicants than spaces typically do rank order on some basis and offer admission from the top down. Waiting lists are established for those who do not quite make the cut. Only four (33%) of the sample institutions currently do this, but, according to the testimony of over half of the program directors, more may be forced to do so as budget cuts impact program size. In one

of the most competitive programs in the sample, the applicant pool exceeds the number of spaces by a factor of four. Most of the candidates not admitted meet the minimum published criteria, but are less well qualified than other candidates according to the faculty who evaluate all applicants. In this program, the decision to admit, according to the statements of the program director, is based upon a general assessment of an applicant's probable ability to teach and the individual's "fit" with the program orientation and philosophy. Thus, GPA, performance on the GRE (this is an MA program), and a general assessment of fitness for teaching determine the overall rank order. While the program does not award official scores, the admitting committee faculty indicated during the interviews that they do use a rough system of weighting the impression made during the interview along with the numerical scores from GRE and GPA. They indicated that an exceptional interview performance can outweigh GPA and GRE to some extent, but candidates with large differences in interview performance and traditional graduate school selection items (GPA and GRE) are not accepted.

Purposes and uses of admissions assessments. These various assessments serve several different purposes, and are discussed below.

Grade point average. The principal purpose of obtaining grade point average, according to the program directors and other individuals involved in the admissions process, is to give a comparable assessment of overall academic ability and subject specific ability. Every program in the sample collects these data and all reported making use of these data in admitting students to their programs. From all the evidence gathered from the sample of institutions in this study, the GPA of teacher education candidates equals or exceeds the median GPA of all students in the major (10 sites) or, at least, the median campus GPA (2 sites). Thus, if overall GPA is a useful indicator of general academic performance and ability (professional and graduate schools have historically used GPA as a key admission criterion), then the individuals admitted to teacher education programs in California have, since 1986, been drawn from the upper half of the college student population. The GPA is seen by the program directors and evaluators of subject matter competence interviewed in this study as the most useful datum in the application process.

Only two programs made any use of the assessments (other than final course grade) done of applicant performance in undergraduate courses. Both programs recruit from their own undergraduate program and have easy access to student course papers and performance portfolios. One program director indicated that the art subject matter assessment included the presentation of a portfolio of prior work and that the music subject matter assessment included an audition.

Subject matter assessment. Eight of the twelve institutions reported that the subject matter assessment of applicants to both multiple and single subject programs is conducted by the appropriate academic department. In the other four institutions, three of them use applicant scores on the National Teacher Examination as evidence of specific subject matter competence and one program uses school of education faculty to assess candidates. No program director or evaluator of subject matter competence made diagnostic use of any of the subject matter assessments of admitted candidates. Once the individual was deemed competent in the subject area, the verification was filed. Only those candidates judged as not meeting the program standards received any feedback and that was in the form of recommendations of courses to take to fill in the gaps in knowledge.

CBEST. Only one program reported that it made any diagnostic use of the CBEST scores of accepted candidates. Credential program faculty were advised by the program director about candidates who scored in the low pass area on the CBEST. The purpose of this information, according to the program director, was to alert faculty to potential problems so faculty could monitor and adjust instruction.

In 80% of the interviews, program directors reported advising applicants who failed the CBEST about remedial course work needed before re-taking the CBEST. The actual scores of candidates who pass the CBEST are not included by any program in this sample in the summary review practices for determining admission. The only issue is whether the applicant passed the CBEST or not.

Document review. All program directors and other faculty involved in making admission decisions reported that the purpose of collecting documents such as letters of recommendation and personal statements is to provide more information about the candidate. They are read by the admitting committee or faculty and evaluative judgments about the candidate's personal fitness, writing ability, and the past work record of the applicant are made. No one reported denying a candidate admission to a credential program on the basis of letters of recommendation or the personal statement. Anecdotal information was given by three program directors about cases where powerful personal statements helped overcome lower GPA's or indifferent academic records. This is particularly so in the cases where the applicant had been away from school for some years or had an unusual school or work history. For the vast majority of applicants to the program in this sample, however, admission documents appear to be pro forma elements to the application process.

Interviews. All the program directors interviewed said that the purpose of interviews was to obtain some sense of the applicant's personal characteristics, make a general judgment of the candidate's oral skills, and hear the applicant's reasons for and expectations about teaching. All the programs use some type of interview rating form and half of the programs use Likert scale assessments of applicant performance in multiple categories. All the forms used encourage the interviewers to make additional comments about the applicant's performance. The interview form is included in the applicant's file and, for those programs that rank-order applicants (33% of sample), performance on the interview is included in the ordering of applicants. For the remainder of the programs, performance on the interview is judged as Pass-Fail with provision for re-interviewing by another faculty. No program reported officially using the information gathered about a candidate in any diagnostic capacity. Half of the program directors reported making unofficial comments about successful applicant behavior in the interview to other faculty. Only three faculty reported making any comments to applicants whose responses in the interview were judged unacceptable. In those cases, the faculty all felt the candidate needed advice about statements made regarding teaching or had personal characteristics that would, in the opinion of the faculty member, make it unlikely that the applicant would be a successful teacher.

While all the individuals interviewed for this study talked about the interviews conducted and indicated they thought the information gained about applicants was valuable, only a few cases of a candidate being denied admission to a credential program solely on the basis of the interview were remembered by any of the persons we interviewed. In those cases, the candidate's performance either appeared psychologically bizarre to the interviewer or the answers given to common questions about classroom actions generated concern regarding the physical safety of children who might have that person as a teacher.

The purpose of requiring exposure to and experience in contemporary schools and classrooms, according to program directors and faculty involved in admission, was more to screen out those whose notions of schooling and teaching were inappropriate than to instruct students in aspects of teaching. If, at the end of the course or experience, the applicant still spoke favorably of teaching, the program directors made no further use of this element of the admissions process. No program, for example, analyzed the applicant's understanding or interpretation of the pre-professional experience in rank ordering candidates for admission.

The summary review procedure only served as a bureaucratic checkpoint and merely ensured that State and institutional requirements were met before offering admission in the

66% of the sample that admitted all qualified candidates. For the 33% of the programs that admitted a fixed number and took the most qualified candidates from the pool of technically qualified, the summary review was a critical element in the admissions process. The program directors of those programs reported spending a significant amount of time and energy reviewing admissions files, weighing evidence, talking with faculty involved in the interviewing process, and balancing the various data points. No program that engaged in such a process made formal use of the rank ordering or individual assessments of admitted candidates after admission to the program.

Technical quality of admissions assessments. Use of overall GPA represents good technical quality and some rigor as these data are based on comparable measurement practices. GPA is commonly used in other graduate and professional programs as an admission criterion. The absence of data regarding the predictive validity of GPA reduces its quality and since 75% of all applicants who complete the admission process are admitted, the technical quality of this assessment point is less important. The CBEST, as a standardized testing instrument, has been carefully reviewed and normed and has withstood several challenges to its validity and reliability.

Subject matter assessment practices vary widely in their technical quality. The National Teacher Examination represents high technical quality as it has been carefully validated by standard psychometric measures. The subject matter assessment practices, as reported by the evaluators of subject matter competence in this study range from the English assessment course which features multiple measures, close proximity to teacher knowledge, and reasonable rigor to the modal procedures which is an interview of the applicant using varying questions by different faculty from the department. While such practices go well beyond transcript review, they do not have the technical quality of the English assessment course. All seven of the CSU evaluators of subject matter competence made reference to planning documents developed to help departments do a better job of determining subject matter competence.

The technical quality of document assessment practices is poor. No programs had formal criteria for judging letters of recommendation or personal statements. According to the statements made by the interviewees, these materials are judged holistically based on the experience of the evaluator in working with credential candidates.

Only one campus reported any formal training in interviewing techniques required of faculty and only two actually converted performance in the interview into some type of point score. The other programs sort the candidate's interview performance into three

categories -- typically identified as Strong, Average, Low or Accept, Possible Accept, and Re-interview. Any likelihood of reliability of interview data across multiple interviewers was due to general conversations among the interviewers. In one small program, it appeared that the principal faculty conducted all the interviews and then met to review all the candidates. Those programs that have capped enrollments (33% of the sample) do attempt to give numerical scores to applicant responses to questions but the evidence is not supportive that such scores represent reliable judgments across applicants. In those small programs where a few faculty conduct all the interviews, there is some inter-rater reliability, but large programs feature many interviewers who may vary in their assessments.

The rigor of evaluating pre-professional experience is very low as it is a binary decision. Either the candidate has or has not completed the required experience. If the candidate has completed the experience and is still interested in teaching, they have met that admissions element.

For those programs that formally have a summary review process, the technical quality of it is moderate. From the statements made by those program directors and involved faculty, the final decision is based on informal criteria and the rank ordering is unrefined and open to personal perceptions. There is an attempt to weight aspects of the admission materials and to make an informed judgment about the likely success of each candidate as a teacher.

Retention in Credential Programs

Range of methods used. Once students are admitted to a professional education program, assessment and screening continues through all course and field work. Those programs that did not require CBEST to be passed or did not require subject matter competence to be verified, apply those screening points prior to the onset of formal student teaching. All the programs in the sample assess candidate knowledge through formal mid-term and final examinations which use a combination of essay and objective questions or essay only questions. The review of the curriculum materials submitted by all the programs reveals that the methods instructors require the development of instructional materials and lesson and unit plans. All students in the sample programs are required (by CTC requirements) to maintain a "B" average and nine of the programs actually dictate that no grade lower than a "B" will count toward certification once a candidate is admitted to teacher education. Performance in fieldwork courses is assessed through observation by university supervisors using adopted observational checklists.

Four programs in the sample make student teaching decisions coterminous with admission and the other eight programs have some type of assessment of performance in the program before a recommendation for advancement to student teaching. The eight programs that do not have student teaching throughout the course of study place candidates in early fieldwork activity that precedes formal student teaching (formal student teaching is here defined as the candidate taking full instructional responsibility for a whole classroom or at least two class periods for the equivalent of a fifteen week semester). These courses generally call for 45 or more hours of observation/participation activities. Those programs that do not initiate student teaching in the first semester of enrollment generally provide instruction in educational foundations, methods of teaching the subject or subjects, and other professionally related courses coupled with a field course that begins with observation activities and culminates in some supervised teaching. The actual teaching done in the early field experience, or informal student teaching, can be linked to planning work done in other classes, but it is designed to give some general sense of the candidate's ability to conduct large group instruction and manage students. In these programs, the student actually teaches one class (secondary) every day for two weeks minimum or one or two subjects (elementary) for two weeks minimum. Supervision is provided by the program and both the master teacher and university supervisor make a formal recommendation about advancement to formal student teaching.

Purposes and uses of retention assessments. University supervisors reported that since candidates see their own evaluations and often are required to sign them, some feedback from this early field experience is obtained. Course grades are filed in all programs in the study and all programs follow the State requirements regarding academic performance. All programs assess the early field experience using a Likert scale of basic competencies. The actual decision to advance a candidate to formal student teaching in the eight programs that have a pre-student teaching field experience is made by the university supervisor with input from the master teacher. The data gathered about students in course grades and the evaluation of the field experiences in all the programs in the study is kept in student files by the program. It is used only when a candidate is in difficulty. The information gathered about students deemed successful in early field experience and course work is, according to the university supervisors interviewed, used informally to make student teaching assignments and is based on faculty and supervisor conversations about the strengths and weaknesses of candidates during this phase of the program. Since so few students are denied advancement to student teaching and those that do tend to be obvious cases of individuals unsuited by temperament for teaching, this informal feedback process was perceived as sufficient by the university supervisors.

All the program directors in the sample indicated that counseling out occurs at the point of recommendation for formal student teaching. The actual numbers could not be verified by the interview teams but, by averaging the percentages given by the program directors and supervisors, 5 - 10 % of admitted candidates leave the program at the point before student teaching commences. Some unknown percentage (note: all of the information about why candidates leave is anecdotal because no program kept accurate statistics of why students leave credential programs) of these students opt out because they realize they are not sufficiently committed to teaching. The majority of the remaining students are counseled out by faculty on the basis of clear difficulty with simple classroom management issues and/or didactic, large group instruction. According to the program directors, one or two students each year per program do not pass the certificate of clearance process which checks for criminal records (California law requires the absence of certain criminal convictions for individuals desiring a teaching credential). Certain types of offenses can require extensive official reviews of fitness for teaching or could prohibit teaching altogether. All teacher education programs advise students at the time of admission regarding this legal requirement but the actual process of fingerprint review can take months which might mean a student would have completed a semester or more of study before obtaining final clearance. An additional few students delay advancement to student teaching due to their inability to pass the CBEST or NTE and, finally, program faculty in the five institutions in large urban areas indicated that a few students each year withdraw for financial reasons.

Technical quality of assessments. The quality of the assessments used for retention in credential programs is moderate. While the grading criteria within credential courses such as psychological foundations of education and methods of teaching are reasonably clear, according to the course materials reviewed, and the competence of the faculty to make those judgments is assumed by virtue of their education and experience, these assessments are less critical to retention in credential programs than performance in the early field experience course. Interviews with program directors revealed that virtually all dismissals from their programs came as a result of poor performance in early fieldwork. The principal means of assessing early field experience in all the programs in the sample, according to the master teachers and university supervisors interviewed, is an observational checklist which covers the basic competencies of teaching. While such a checklist is completed for each candidate, interview data suggests that the university supervisor and the master teacher make holistic judgments about their candidate rather than making individual judgments on the various criteria and summing those scores to reach a final judgment.

Formal Student Teaching

Range of methods used. When formal student teaching begins, the length of time, the number of periods, and the ordering of the student teaching experience are typically spelled out in a handbook given to all the participants (student teacher, master teacher(s), and university supervisor). All the programs in this study have such a handbook and many of them are quite elaborate and contain helpful hints about clinical supervision, timelines for student teachers to follow, bulleted reviews of basic classroom management and lesson planning, and a series of expectations for all the people in the student teaching triad -- student teacher, master teacher, university supervisor. These handbooks also include the various evaluation forms used by the master teacher and university supervisor to provide part of the formal record of successful student teaching. While other data is collected about the performance of the student teacher (e.g., journals, lesson plans, videotapes), the official assessment instrument and determinant of a recommendation for a credential is the summative assessment checklist.

While the length of formal student teaching is defined by CTC regulations, there is some variation in how institutions meet that requirement. The one program that features student teaching throughout the program tends toward part day teaching assignments coupled with classes in the later afternoon. In the case of single subject candidates, they typically observe and then teach in two periods over the course of two semesters while the multiple subject candidates observe and teach a variety of subject lessons during the morning in the first semester and cover the remaining subject areas in the second semester. In all cases, the key moment in student teaching occurs at "turnover" when the student teacher is fully in charge of the elementary class for the entire day or when the secondary level student teacher is fully in charge of all classes assigned.

It is probable that much subject matter pedagogical discussion goes on among the triad members (student teacher, university supervisor, master teacher), but there is no official verification of subject specific competence in the institutional forms. There is a question about whether the candidate knows the subject matter, but only one of the forms asks specific questions regarding teaching strategies that are particularly important to a discipline or subject, a question on effective safety procedures in the laboratory. Based on the interview data with master teachers and university supervisors, many informal discussions during student teaching center on such matters but are not included in the formal formative or summative assessment of student teaching.

Two programs require students to keep journals and these are read and commented upon by university supervisors, but again, in reviewing documents and handbooks, there does not appear to be a linkage to the final recommendation. They are used to help the university supervisor gain a richer understanding of teacher thinking and growth during student teaching.

The student teaching course does not have much assigned work beyond what is required in the handbook or what is required by the master teacher. Mostly, student teaching consists of taking over an increasing amount of the teaching day if one is an elementary credential candidate or taking over two or three classes at the junior or senior high school level. All candidates in California must have a two level experience - K-3 and 4-6 for multiple subject candidates and 7-9 and 10-12 for single subject candidates. Additionally, all student teaching must take place in classrooms where at least 25% of the students are of a different ethnic, racial, or linguistic group from the student teacher. Most student teachers appear to teach in average level classes rather than advanced or remedial ones. Most student teachers work in schools that are urban or suburban with some students teaching in inner city schools. The smallest number of student teachers appears to be those teaching in rural settings.

Purposes and uses of assessment. The university supervisors in this sample of programs sees a student teacher once every two weeks, on average. Two of the small credential programs claim they provide weekly supervision and half of the programs also have a weekly campus seminar that accompanies student teaching. The evidence garnered from master teachers suggests that the "every two weeks" rule is an average and that some student teachers get more supervision than others. The length of the university supervisor's stay also appears to vary individually, according to the master teachers. Some university supervisors reported staying for two or more hours per visit in order to see a whole class period plus conference time. Others reported varying the time to account for individual situations.

All university supervisors reported that the primary purpose of the visits was to ensure that the candidate was meeting minimal standards of teacher behavior and making improvement on the criteria listed in the student teaching handbook. The one program that required student teaching throughout the program developed a clear set of written expectations for candidate behavior at selected points in the program. The remaining programs are less clear on what competencies might occur when or whether beginning teacher competencies develop simultaneously during student teaching.
In addition to checking on student teaching performance, all the university supervisors

reported that conferring with the master teacher and the school principal was a secondary purpose of the field visits. Getting useful feedback from professionals who are on site every day was perceived as very useful. Formal three-way or four-way conferences were required once or twice during student teaching in all the programs reviewed. The actual content of these conferences was left to the discretion of the university supervisor.

From the interview data with university supervisors and master teachers, the principal use of the information collected during student teaching is to buttress the decision to pass student teaching and recommend the student for a credential. The amount of written information accumulated about the candidate is a function of his or her performance. The poorer the performance, the greater the amount of information collected. The reason, according to the interview data, is that student challenges to faculty decisions require documentation. Successful student teaching merely requires a passing grade and a letter of recommendation from the master teacher.

All the university supervisors reported making use of their field notes as formative evaluations. The supervisors from the two programs that required journal keeping also reported making extensive use of those documents through a running commentary they made about the journal entries. The master teachers reported a greater use of oral communication.

Technical quality. The handbook forms that are used during student teaching are focussed exclusively on generic pedagogical principles that mirror the CTC competencies and the process-product research literature. None of these forms have precise descriptors attached to each performance level. One program requires student teachers to develop portfolios of work which are commented on by master teachers and graded by university supervisors but it is not clear in the student teaching handbook whether such work affects the summative recommendation about student teaching. One campus has an elaborate form for both its basic credential programs that requires formal sign-off by the university supervisor and the master teacher on each of the ten pedagogical competencies established by the CTC. This form is in addition to its own summative evaluation of student teaching.

In reviewing all the forms used to assess student teaching from the study sample, the number of individual behaviors or competencies evaluated ranges from 15 to 65 with the average number in the mid-20's (Table 2.4). All of the forms look at some type of planning skills, both lesson and unit, classroom management, instruction or presentation skills , interpersonal and/or professional skills, and some forms address evaluation of student work skills, room arrangements, small group/large group work and materials developed. The lack

TABLE 2.4

DISTRIBUTION OF TEACHER COMPETENCIES BY CATEGORY

Categories	Selected Programs from Study Sample (7 of 12)						
	A	B	C	D	E	F	G
Planning	2	3	4	4	4	7	23
Instruction	9	6	6	8	8	13	11
Evaluation	0	0	0	0	0	0	3
Interpersonal Skills	2	0	0	0	0	0	10
Professional Behavior	2	5	4	5	19	12	8
Materials Use and Development	0	0	4	0	0	0	0
Room Arrangements	0	0	2	0	0	0	0
Classroom Management	0	4	0	6	0	10	3
Diagnosis	0	0	0	3	0	0	7
Teaching Mechanics	0	0	0	0	0	5	0
CTC Standards	0	0	0	0	10	0	0
TOTALS	15	18	20	26	41	47	65

of a common language for describing the various aspects of teaching is evident in the differing names given to clusters of skills. In addition, some programs place greater emphasis on some aspects of teaching than others (e.g., 19 professional behavior elements versus 2 elements). It is evident, however, that all programs put emphasis on planning, instruction, and some set of behaviors involving classroom management.

Since the master teacher plays such a critical role in the formative evaluation of the student teacher by virtue of daily interactions and because the current CTC standards require that the master teacher make a judgment about the competence of the student teacher, some analysis of the training and support given to master teachers is important to understanding the technical quality of assessment during student teaching.

In analyzing the responses of all the individuals involved in student teaching, it is evident that the selection of the master teacher, while a critical element in a successful student teaching experience, is not easily explained. All the institutions in this study indicated that they select only those individuals who have experience and who are considered well qualified to supervise a beginning teacher. Evidence from interviews with master teachers in at least four of the programs suggests that is not always the case, but that institutions try to get good role models as master teachers. One institution has developed a clear set of application materials in conjunction with the local districts and limits the pool of master teachers. Other programs appear to have more informal arrangements and build up a cadre of good master teachers through personal contacts. One cautionary note is that the pool of available, trained master teachers really needs to be three times the number of candidates in the pipeline to account for not over using the same schools and teachers. Thus, a program that admits 100 students needs about 300 approved master teachers who meet the CTC requirements noted above and who teach in schools within easy commuting distance for students and university supervisors. No program in this study limits the number of students it admits to the number of high quality master teachers available although some of the small programs in densely populated areas may achieve that state naturally.

Given these concerns and the lack of recent training for master teachers provided by the teacher education programs in this sample, the ability of the master teachers to utilize the assessment forms provided by the teacher education program appears problematic. While all of the master teachers interviewed felt capable of making the overall decision about recommending a candidate for a credential, over a third felt uncertain about the meanings or intentions of the sub-parts of the summative assessment form they were to use.

Recommendation for the Credential

Range of methods used. For all but one of the programs, the recommendation for the credential is contained in the passing of all required courses, the recommendation received from the university supervisor and master teacher, and verification by the program director or credential analyst that all State required matters have been met. At this point, the candidate files the necessary forms for the credential through the university, establishes a placement file, and seeks employment. The one program with the portfolio requirement also requires that the portfolio be successfully evaluated. In all programs, failure to complete student teaching with an approval prevents a credential recommendation.

Purpose and use of assessments. In analyzing the program documents, no program requires a minimum score on the individual competencies listed on the summative student teaching forms to pass student teaching. That is, none of the programs made specific statements that failure to obtain a specific rating on certain competencies would result in a failure to compete student teaching. Recommendation for a credential constitutes a global assessment of the basic competence of the teacher. While everyone uses some type of competency based form complete with a 3 to 5 point Likert scale or some form of individuating among the candidates, the language used by program faculty during the interviews to describe the difference between a candidate who is successful and one who is not suggests the decision turns on an overall assessment rather than on one or two key competencies. That is, good candidates are generally good and poor candidates are generally poor. Thus, it is difficult to isolate a single predictive variable that might explain who should or should not get a credential.

Technical quality. The technical quality of the assessment practices during student teaching is low due to the absence of clearly defined descriptors of levels of competence in the observational checklists used by university supervisors and master teachers. Additionally, none of the programs in this sample had current formal training activities for those individuals asked to fill out the forms. The complexity of the behaviors embedded in the official forms and the absence of formal training in their use raise questions about inter-rater reliability and equity across programs which reduces the quality of the assessments.

Post Credential Support

Range of methods used. All credential programs are required by CTC regulations to engage in regular evaluation of their programs by program participants and local educators. These evaluations historically have been some form of a questionnaire which addressed how

well the graduates were prepared to function as classroom teachers. All of the programs in this sample reported using a written questionnaire to obtain feedback. Three of the programs in the sample have received grants from State agencies fostering new teacher support to provide additional services to first year teachers and to gain additional feedback from those teachers to improve their pre-service program. The nine programs which have not received outside grants have attempted more informal arrangements with graduates and claim to benefit from feedback from their students who later enroll in advanced degree programs or advanced credential programs. No program reported formal interactions with employing school districts regarding the first year performance of their graduates in relation to district expectations for first year teachers.

Purposes and uses of assessments. The basic purpose of post credentialing interaction with graduates of credential programs is for the program to get feedback about the appropriateness of its curriculum and the general quality of its preparation. Much of this activity is informal rather than formal. The data collected from graduate questionnaires is used in faculty discussions of the program, according to the program directors.

Technical quality. The questionnaires used in the sample programs are standard opinion instruments that contain questions on all or part of the credential program. They are based on the particular curriculum in each program and are, therefore, hard to compare. According to the program directors, the data are summed in the form of percentages and faculty draw conclusions based on the perceived value of the responses. No program director indicated that the questionnaires had been tested for validity and reliability, although they appear to have face validity in that they contain questions about the course work and field experiences in each program. All other efforts to assess the performance of credential graduates in first year teaching positions are informal conversations and cannot, therefore, be judged on technical merits.

Focus of IHE Assessments

This section of the report reviews the types of assessments engaged in by pre-service teacher education programs from the perspective of the "why" of the assessment rather than the "what" of the assessment. The underlying knowledge, skills, or abilities that a teacher must possess should be embedded in the assessment practices and instruments that faculty and supervisors use to make admission, retention, and recommendation decisions.

Subject-matter

Although the undergraduate programs engage in repeated assessment of credential candidates through course examinations and the like, the subject matter assessment at the point of admission to credential programs makes no use of that information. Transcript evaluation by the faculty represents the primary determination of performance in previous course work. The emphasis in transcript review is ensuring coverage of the broad content areas and the matching of courses taken to the appropriate State framework. The interviews that serve as the dominant means by which academic departments assess competence in the subject matter also emphasize breadth over depth.

The more authentic assessment of subject matter knowledge is understood by academic department faculty and effective plans have been developed but await implementation. It should be noted that the faculty interviewed in this study asserted that they can tell a great deal about a candidate from a careful reading of a college transcript. That information coupled with a thoughtful interview conducted by an experienced interviewer can produce a detailed assessment. Once subject matter competence has been assessed at admission, it is no longer reviewed in any formal way by any of the programs in this sample.

General Pedagogy

Once admitted to the program, the focus of assessments turns to performance in actual classrooms. While all programs require students to maintain a certain GPA and a few students run afoul of that rule, the evident interest is in how the candidate performs with real students in real classrooms. Moreover, there appears to be a heavy emphasis on the performance aspect of teaching as opposed to the planning aspect of teaching. That is, all the examples of students pulled from student teaching or made to repeat student teaching turned on lack of success in the performing part of teaching, i.e.,classroom management, classroom presentations, and classroom routines. There was little clear evidence that candidates fail student teaching on the basis of poor lesson plans or other non-performance aspect of teaching if there is no other concern. Criticisms of such students abound in the files, but, with some exceptions, candidates who can handle the normal classroom routines are rarely denied certification.

In sum, it appears that the key sets of competencies are:

- classroom management sufficient to allow order and typical classroom activities to occur;
- organizational skills sufficient to handle normal classroom and school requirements (attendance, etc.);
- reasonable lesson and unit plans such that the required curriculum is covered in the time allotted;
- reasonable knowledge of subject matter so that egregious errors are not made and parental complaints are avoided;
- standard work behavior - punctual, clean, sociable, able to interact with other adults in the school and the community.

Content Pedagogy

None of the programs in this sample formally assess subject specific pedagogy. While data from interviews with master teachers and university supervisors suggest that this area of beginning teacher competence is addressed in informal discussions, there is no evidence that teaching competence within subject areas is formally evaluated during student teaching. One program had one item related to lab safety on its student teaching evaluation form, but all other forms focussed exclusively on competencies and behaviors that were not subject specific. The one program that required a portfolio did afford its students an opportunity to write about their subject matter, but the assessment of it was not subject specific.

Other Constructs

Basic skills proficiency. From the data gathered in this study, it appears that institutions of higher education assess basic academic skills reasonably well. With the CBEST, writing samples, interview data, and above the median GPA required for admission coupled with the evidence of repeated assessments of candidate basic skills ability in college course work, additional assessments seem unwarranted. The continuing difficulty of some groups of students to pass the CBEST suggests some efforts are needed to aid students weak in basic skills early in their college work, but this is not the particular responsibility of teacher education programs.

Personal attributes. Every program conducted interviews of candidates and all programs focussed on exploring the stability of the candidate's desire to teach and the experiential basis for that career decision. The admitting committees claimed to be looking for a set of personality traits and attitudes that suggest the candidate would be an effective teacher. These appear to be grouped around a social commitment to helping others, a multicultural orientation, and a concern for social improvement and justice. The perceptions of the faculty in the one program with extensive psychological screening were that such assessments did predict success in the program and the classroom; however, no data had been collected to support that perception. In reviewing the statements made in the interviews by faculty involved in the interview process, the notion of screening out those who are clearly unsuited for teaching by virtue of obvious negative personality traits or unequivocal statements about children or youth that are inappropriate is clear. When pressed about their ability to assess candidates on personality traits, the faculty in this sample only could give specific examples of extreme cases.

Programmatic needs. In those few programs that have capped enrollment and have a larger pool of applicants than spaces, this administrative decision to limit enrollments has the net effect of being a powerful screening tool. It is in these programs that the complex weighing of information is most evident. For these programs, GPA and subject matter verification are not enough to warrant admission. In cases where an otherwise capable candidate is deemed lacking in these qualities, another candidate might well be chosen.

One other characteristic appears to influence admissions decisions. Most admissions decisions are based on brief interactions and data that are not clear explanatory variables for successful teaching. Most faculty interviewed about admissions decisions seem prepared to give candidates a chance to prove themselves as teachers and often felt strongly that they should give the marginal candidate the benefit of the doubt. One respondent argued, in support of the notion of caution, that very bright students make poor teachers so that over-reliance on GPA was a poor idea. Others felt that general intelligence was the best predictor of success and were not concerned with notions of previous experience or completion of waiver programs. Given the undergirding uncertainty about predicting who will be an excellent teacher, it is not surprising that programs err on the side of inclusion rather than exclusion.

Differential Impact of Assessment Practices

It appears that the GPA requirement, particularly the post-1985 requirement of the CSU, has provided a clear pre-screening device since all the advisement materials reviewed clearly state the system-wide requirement. Since most students know their GPA, it is likely that some students make no further inquiry and do not pursue applications. This phenomenon may explain the difficulty of determining whether such requirements have an adverse impact on under-represented groups. There is insufficient evidence from data on who is denied admission to teacher education to argue that the GPA requirement adversely impacts under-represented groups. The comments from program directors and others involved in the admissions process suggest that individuals from under-represented groups comprise less than 25% of the applicant pool in this sample of credential programs and, thus, are not denied admission in significantly higher percentages than other candidates. Increasing the number of applicants from under-represented groups appeared to be the larger problem reported by program directors. Most faculty indicated a willingness to interpret GPA in light of personal experiences and give more weight to recent performance than earlier academic work. Official policy in the CSU, according to program directors, permits up to 15% of the admitted group to fall below the median GPA requirement. That is, every year each teacher education program could admit a percentage of its candidates from the group whose GPA was below the median for that major and that campus if the faculty felt there were other characteristics that suggested likely success as a teacher. While the data are anecdotal rather than numeric, none of the program directors in the CSU campuses indicated that they had fully used their 15% exceptions clause on behalf of under-represented groups. The low number of under-represented persons entering teacher education programs does not appear to be directly linked to the raised GPA standards.

The use of undergraduate GPA is pervasive and is linked to admission to the college or university when post-baccalaureate students are applying to teacher education programs. Since many colleges have fairly liberal post-baccalaureate admission policies, this is rarely a step that screens out candidates. A few programs require additional tests (GRE) since their credential programs are linked to a graduate degree. Eighty-three percent (10 of 12) of the sample institutions offer only the credential rather than a combined credential and graduate degree program.

The pattern of adding courses to the pre-requisites for admission, coupled with the need for late-deciders on teaching to take the NTE, was suggested by faculty involved in the admissions process as an adverse impact on under-represented groups who may have difficulty with standardized tests. The dilemma of poorer applicants who cannot afford the

additional time without an income was mentioned by three program directors as an adverse impact of assessment practices. They were unable to produce actual percentages, but remembered several cases from recent years.

Five of the institutions reported that the CBEST appears to be the barrier to under-represented participation in teacher education. Four of the institutions reported that they had too few under-represented applicants to determine the impact of the CBEST. Since the inception of the CBEST, the passing scores of under-represented groups have been a pressing concern to program directors and admission committees. Some campuses report providing workshops for students needing help but the CBEST requirement remains as a significant challenge to teaching for minorities and under-represented groups.

With the exception of the one program with the complex array of psychological tests, the data suggest that programs do not deny candidates solely on the basis of poor interview performance unless the behavior of the candidate is clearly unacceptable (to the level of antisocial/nonsocial behavior evident to an untrained observer). As with the letters of recommendation, this appears to be a pro forma activity that actually screens out very few applicants. There is anecdotal evidence that even individuals who give pause to the interviewers are not denied or delayed for reason of interview performance but rather some other reason is given. This, according to the comments made by individuals involved in admissions work, stems from a worry that interviews are weak indicators of potential for teaching and would be deemed insufficient evidence if the applicant protested the decision. Given the absence of training for interviewers and the lack of inter-rater reliability controls, such caution appears warranted.

There is little evidence from the interviews that anyone is denied admission on the basis of prior experience with children or youth. Applicants are delayed until such time as they have had recent experience with children or youth in a school setting. The institutions that have more closely defined what is meant by experience with children and youth appear to be trying to ensure that applicants have a realistic understanding of the work life of teachers. The evaluation of the experience and what it actually taught the candidate is difficult to surmise from the data collected. Recent, verified experience is seen as a general good and programs assume all candidates have automatically derived sensible and appropriate experiences from the exposure.

The issue of selectivity of teacher education programs is not a simple matter to determine. The widespread publicity about minimum GPA standards screens out some applicants from the start. The requirement that CBEST be passed before admission may

ward off those who believe they might not pass the CBEST. The increased scrutiny of prior field experience (e.g., requiring actual participation in schools) may dissuade some candidates from pursuing application or at least cause a delay in admission and increase the chance of economic hardship for these candidates needing to work full-time. The increased attention to subject matter competence, either through passage of the NTE or completion of a "waiver" program, may also hinder some applications if the student has sought informal assessment from an academic department prior to submitting an application.

Apart from these caveats, it appears that with few exceptions, 75% of applicants to teacher education programs are eligible for admission. Of the 25% that are not admitted or admissible, about 15% are advised to re-apply when they have completed some element of the admissions process. Typically, this would be the pre-professional field experience or subject matter verification. The remaining 10% are denied, most commonly on the basis of GPA or failure to pass the CBEST.

Admission procedures for teacher education programs are as complex, if not more so, than procedures for admission to other professional programs. While teacher education programs do not uniformly require such aptitude tests as the GRE, GMAT, LSAT, MCAT or Miller's Analogy, they do seek multiple sources of information about the candidate from a variety of individuals and do attempt to show a direct link between the knowledge, skills, and experience of the applicant and the work of a classroom teacher. The technical level of these screening activities is not high, but they are not appreciably lower than those used in other academic admissions.

The general thrust of all screening after admission has a flavor of identifying the failing candidate and promoting those who are making reasonable progress toward demonstrating that they can be trusted with a classroom. This emphasizes management issues more heavily than any other issue. The forms used to evaluate the field performance of the candidates are focussed on generic pedagogy and basic professional behavior. None of the programs in the sample used subject specific evaluation forms. One of the programs that features heavy field work throughout the program has developed an evaluation form that is developmental in expectation. As students gain more skill and knowledge during the year, the number of competence areas marked "not observed" is expected to drop and the description of what constitutes acceptable performance rises. For most of the programs in this study, the decision to advance a candidate to formal student teaching is based on the candidate having met the State mandated requirements (CBEST, NTE or subject matter competence, certificate of clearance), successful completion of all required course work to date, and acceptable performance on the field work required by the program. In

interviewing university supervisors and master teachers, the descriptions of what constituted unacceptable work turned on very basic notions of professional behavior (e.g., punctual, properly attired, maintains adult behavior, dependable) and fundamental classroom management and instruction skills (e.g., attendance, gives directions, keeps students generally on task, calls classes to attention and dismisses classes). Although students in all the programs in this study are required to prepare lesson and unit plans as a part of the required methods courses and a number of these courses feature some form of micro-teaching or practice presentations, the link between such courses and fieldwork is not overtly set out in any program. It does appear that basic congruence exists, given the fairly constricted curriculum available under current Title V regulations. However, a number of programs appear to permit students to take the pre-student teaching courses in different order, at different times, and some variations seem to exist across various sections of the courses. All these variations add to the difficulty of assuring that each candidate has been exposed to similar instruction and engaged in similar controlled activities prior to being assessed for readiness to student teach.

Costs of IHE Assessments

This section of the chapter addresses the costs of conducting existing assessments. Calculating such costs was very difficult as few institutions compute all the faculty and staff time for this aspect of credential programs separately. Additionally, the three systems of higher education represented in this study calculate employee time differently. The closest common measurement unit among the various programs reviewed in this study was FTE or full time equivalent employee. It is a fairly common means of expressing work load and made it possible to convert actual work hours to numbers of employees. Where interviewers were given data expressed in clock hours, those data were converted to FTE using a standard forty hour work week.

Admission Costs

The time and, therefore, costs involved in assessing GPA is minimal since the data are generally contained on the transcript. In a few cases, according to the program directors, the GPA must be manually calculated by an admissions clerk since a few colleges use descriptive reports or do not calculate total GPA. A few program directors try to assess the number and type of courses taken for Credit/No Credit as such courses can give a different view of the candidate while not altering the actual GPA.

The interview portion of the admissions process takes the greatest amount of faculty time. The largest program in this study has more than 500 applicants to interview each semester. If the interview includes some form of subject matter assessment, that assessment must be evaluated separately and, often, by an academic specialist. This can mean two interviews per candidate. If one assumes that interviewing a candidate about general fitness for teaching takes 30 minutes, a program with 500 applicants would need one FTE faculty member just to conduct the interviews. This assumes that the interviews can be spread across one semester, but, in reality, the admissions process must be completed in time for early registration and schedule refinement. Thus, most large programs compress the interview process into a two or three week period. If one adds in the number of interviews and assessments planned to determine subject matter competence, the allocated FTEF - full time equivalent faculty (since it is only faculty that can make such determinations, not staff) will equal 2. The program in this study that had made the most progress in innovative assessments of candidate's subject matter knowledge did so by making the assessment process an actual university course for which the instructor received academic support and workload credit. The faculty who assess subject matter competence in the large CSU programs pointed out the current lack of workload credit given for subject matter assessments unless they are embedded in a credit bearing course.

Combining the time involved in interviewing and processing the application, according to the data received from all those involved in admissions work in the study sample and creating an average, each applicant takes up between 1.5 and 2.0 hours from initial contact to enrollment in the program. It is important to note that these figures do not take into account all the time spent in pre-application questions or in advising non-students who wish to obtain a credential.

Once admitted to the program, the costs of retention and recommendation for a credential are covered in the faculty assigned to instruct the student or to direct the program. While all program directors claimed time for handling problems with students, counseling potential dropouts, and conferring with school districts and master teachers, none thought such issues took more than a third of their time. The supervision of student teachers is the primary function of the faculty member assigned to that course. In the CSU programs, the system-wide formula is 25:1. That is, a full time employee (FTE) must supervise 25 student teachers in a semester. The workload patterns of all the CSU programs in this study revealed a pattern of giving some supervision to many faculty rather than concentrating supervision with one or two individuals. Thus, beyond the direct costs of paying mileage for the field supervisor to travel to the site and the lost time to the faculty

member of that travel, the costs of retention and recommendation are part of a normal workload for a faculty member.

In sum, the costs of all assessments in basic credential programs reviewed in this study were only an issue in the larger credential programs where subject matter assessment activities required substantial investment in faculty time outside of instructional time. The other costs are covered through administrative support and instructional allocations.

Contributions of the IHE Assessments to Credentialing Recommendations

Almost all of the information gathered about candidates in pre-service teacher education programs stays with the program. Very little is transmitted to employers. Part of the reason is that the placement file is controlled by the candidate and only information approved by the candidate is entered. Typically, this will include the college transcript(s), letters of recommendation including the critical letter(s) from the master teacher(s), and a resume of prior work and volunteer experience. All other information generated about the student during the pre-service preparation program remains at the university. From the interview data, it was learned that an employer might call a professor or university supervisor for additional insights and information. The formal record diminishes when the credential candidate becomes a graduate. The data also suggests that institutions make little use of the assessments collected about candidates in any formative way. No program routinely analyzes student performance in student teaching to revamp courses or expectations. While individual faculty may be aware of trends in candidate performance, such information does not appear to be part of regular program evaluation and development activities.

Program directors noted that information about students is privileged data and cannot be forwarded without permission of the candidate which suggests it is difficult to have the materials accumulated by the teacher education program used by others. Moreover, the information on admissions patterns indicates that candidates in California frequently attend two or more colleges to obtain the baccalaureate degree and attend yet another institution to earn the credential. This profusion of institutions adds to the difficulty of creating a uniform flow of information that builds from the undergraduate years into the early years of teaching.

The impact on credentialing decisions of the reforms in pre-service teacher education programs appear strongest and clearest at the point of admissions where the insistence on screening for basic educational skills and the median or higher grade point average have

had significant effect. The extensive use of qualitative information (interviews, letters of recommendation, etc.) does not seem to have similar weight except in one specialized program. The existing plans for assessing subject matter competence appear comprehensive and thoughtful but are not routinely employed and have not been evaluated for technical quality.

Once admitted to a teacher education program, most students complete the sequence of courses and field experiences. Dropping out of the program appears to turn on financial difficulties and diminished interest in teaching as much as outright failure to meet expectations. The expectations for student teachers are not unsurprising although the technical quality of the instruments used to determine threshold skills is low and the training in their use not uniform. There is increased interest in use of portfolios and other means of determining student teacher competence, but these remain experimental and are not yet connected to the final decision to recommend a candidate or not. That decision still remains in the realm of Likert scale instruments that focus on generic pedagogical skills as defined by the Commission on Teacher Credentialing and the research base of the last ten years.

Summary and Conclusions

Key Admissions Elements

All institutions with approved teacher education programs are required (since 1986) to demonstrate that the average GPA of admitted students exceeds the average GPA of some comparable group of students. All CSU campuses have been required since 1985 to limit admission to teacher education to those candidates whose overall undergraduate GPA exceeds the median GPA for students in the same major. There is insufficient evidence from data on who is denied admission to teacher education to argue that the GPA requirement adversely impacts under-represented groups. The evidence suggests that too few under-represented individuals apply at all to the programs.

All programs in the sample gathered information about subject specific knowledge as well as general academic performance. In California, all teachers must demonstrate competence in the subject they will teach and do so in one of two ways (NTE examination or completion of approved program of study). In the CSU, the determination of subject matter competence, whether by examination or by completion of an approved program of study, is to be verified by the academic department on campus.

All programs must require applicants to take the CBEST as a part of the admissions process and 60% of the sample requires a passing score on the CBEST before admission to teacher education. The CBEST appears to be the barrier to under-represented participation in teacher education. Passing scores are set by the State and there are no exceptions permitted.

Letters of recommendation and personal statements are used in the admissions process to provide some judgments about the writing ability and the past work record and personal qualities of the applicant. They do not seem to be critical variables and no program director or admission director reported denying a candidate admission solely on the basis of letters of recommendation or the personal statement.

The data suggest that campuses do not deny candidates solely on the basis of poor interview performance unless the behavior of the candidate is clearly unacceptable (to the level of antisocial/nonsocial behavior evident to an untrained observer). Additionally, there appears to be little evidence that anyone is denied admission on the basis of insufficient or inappropriate experience with children or youth. Applicants appear to be delayed until such time as they have had recent experience with children or youth in a school setting. 75% of applicants to teacher education programs are eligible for admission. Of the 25% that are not admitted or admissible, about 15% are advised to re-apply when they have completed some element of the admissions process. Typically, this would be the pre-professional field experience or subject matter verification. The remaining 10% are denied, most commonly on the basis of GPA or failure to pass the CBEST.

Key Retention and Recommendation Elements

The general thrust of all screening of candidates after admission is focussed on deciding who is not failing and, secondarily, who is making reasonable progress toward demonstrating that the candidate can be trusted in a classroom. This emphasizes management issues more heavily than any other issue. Moreover, there appears to be a heavy emphasis on the performance aspect of teaching as opposed to the planning aspect of teaching.

The programs in this study make numerous efforts to assess candidate's generic pedagogical ability. The technical quality of these assessments is not high and they appear to be open to inter-reliability concerns as a result, but the competencies noted in the assessment forms are drawn from the research literature on effective teaching and need not be modified. Where additional work needs to be done is in the development of operational

definitions of what constitutes appropriate examples of an "appropriate classroom environment" or "instructional strategies, activities, or materials that are appropriate to a particular class or grade level." In addition to this development, all individuals who use such forms need training in their use. If teacher education programs wish to move away from global assessments of generic pedagogy, they will need to raise the technical quality of their assessments and make clear that certain competencies must be met at a specific level for certification to follow.

While the interviews with university supervisors and master teachers indicate it is probable that much subject matter pedagogical discussion goes on among the triad members, there is no official verification of subject specific competence in the institutional forms. The determination of subject matter competence takes place well before the on-set of fieldwork and is not integrated with the recommendation for a credential. Thus, the recommendation for a credential is a global assessment of the basic competence of the teacher. Candidates who can handle the normal classroom routines are rarely denied certification.

In sum, it appears that the key sets of competencies that must be displayed if a candidate is to be recommended for a credential are:

- classroom management sufficient to allow order and typical classroom activities to occur;
- organizational skills sufficient to handle normal classroom and school requirements (attendance, etc);
- reasonable lesson and unit plans such that the required curriculum is covered in the time allotted;
- reasonable knowledge of subject matter so that egregious errors are not made and parental complaints are avoided;
- standard work behavior - punctual, clean, sociable, able to interact with other adults in school and community.

CHAPTER 3:

LOCAL EDUCATION AGENCY ASSESSMENT PRACTICES

CHAPTER 3:

LOCAL EDUCATION AGENCY ASSESSMENT PRACTICES

The focus of this substudy is on local education agency (LEA) assessment practices. It was designed to capture the range of ways in which new teachers are assessed in local public school districts in California. In most public school districts in California, prospective and beginning teachers are typically evaluated for the purposes of hiring and of making decisions about continued employment and tenure within a district. In addition, as more districts provide beginning teachers with support, formative assessments of beginning teachers may be conducted to determine the types of support needed and/or to monitor teacher progress and development.

This chapter discusses the major findings from our study of the processes and criteria used to evaluate beginning teachers at three major assessment points: hiring, continued employment and tenure, and support decisions. The data reported were collected during site visits to 12 public school districts from May 9 through June 27, 1991. The chapter includes eight major sections. In the first section the sample and methodology employed in this substudy are briefly described. The second section describes the context that shapes current LEA assessment practices. An overview of the assessment practices at each major assessment point is provided in section three followed by a discussion of the focus of assessments that are critical to decisions made about hiring, award or denial of permanent contracts, and support for beginning teachers. In the three sections that follow, brief discussions of the differential impact of assessments, costs of assessments, and contribution of assessments to staff development decisions are provided. The last section summarizes the major findings and conclusions about LEA assessment practices.

Description of Sample and Methods

Sample Description

To study the range of LEA assessment practices currently in use in California, 12 public school districts were purposively selected to capture a range of variation in district settings and conditions, while simultaneously maximizing the diversity in assessment practices. Although the purpose was to capture the range of assessment practices rather

than to be representative of the population, the sampling plan attempted to represent variability in district size, organizational structure of districts (i.e., elementary, high school, or unified district), geographic location, and community type served. Table 3.1 shows distribution of sample districts according to geographic location and size, and geographic location and type of community served. As a measure of size, student enrollment is used. To place our sample of districts within a state perspective, of the over 1,000 public school districts in California, the median district has about 1,300 students enrolled. Moreover, 50% of the students in California are enrolled in one of the 65 districts in the state with 15,675 or more students. The 10 largest districts enroll about 25% of the students in California. Cast against this background, our sample of 12 districts includes one district with student enrollments less than 1,300 (i.e., below the median), seven districts with enrollments of more than 15,000, and four districts with enrollments somewhere in-between 1,300 and 15,000 students. We also have three of the State's 10 largest districts in our sample.

As indicated in Table 3.1, the sample sites are almost evenly distributed between Northern and Southern California with one district in the Central Valley area. We also have a range of urban, rural and suburban sites, though considerations of time constraints and costs led us not to include many districts in very remote areas. However, at least three districts are located in isolated small towns at least one hour away from a major metropolitan area.

In terms of organizational structure, most of the school districts in the state of California are elementary districts, but about 60% of the students in California are enrolled in unified school districts. In our sample, three-quarters of the districts are unified districts (9 of 12), two are elementary school districts and one is a high school district.

In order to ensure that the sample captured a range of assessment practices, other factors thought to affect the ways in which districts evaluate new teachers were also considered. Our sampling plan also included districts that varied along several other dimensions: use of innovative as well as typical methods of teacher evaluation, ability to be highly selective of teachers, degree of difficulty staffing schools (which may result in less stringent criteria), and rapidity of student population growth (which could affect the amount of attention districts are able to devote to teacher assessment).

TABLE 3.1

DISTRICT SAMPLE BY GEOGRAPHIC LOCATION AND SIZE AND COMMUNITY TYPE
(N = 12)

GEOGRAPHIC LOCATION BY SIZE

DISTRICT SIZE (Student Enrollment)	GEOGRAPHIC LOCATION		
	Northern	Southern	Central
Very Small (less than 1,300)	1	0	0
Small (1,500-15,500)	1	2	1
Medium (15,500-39,000)	2	3	0
Large (More than 39,000)	1	1	0

GEOGRAPHIC LOCATION BY COMMUNITY TYPE

COMMUNITY TYPE	GEOGRAPHIC LOCATION		
	Northern	Southern	Central
Urban	2	3	0
Suburban	2	2	0
Rural	1	1	1

Methodology

The study design called for site visits to each of the sample districts to gather information about how beginning teachers are evaluated for hiring, continued employment, tenure and support. Each district was first invited to participate through a phone call made in April to the superintendent by California Department of Education staff. All districts selected agreed to participate. Site visits were conducted between May and June.

One of the main methods of collecting data for the study was through interviews with persons who are involved in beginning teacher assessments. Because the people who participate in continued employment and support decisions varied considerably from district to district, we identified the individuals to interview more by function than role. Table 3.2 shows the distribution of people interviewed according to their role and type of community served. Therefore, the number and types of individuals interviewed varied from district to district. In the 12 districts, we interviewed a total of 152 people, including 108 school-level administrators and staff, and 44 district-level administrators and staff. People interviewed about support included 35 mentors and 13 district staff involved with beginning teacher support and staff development.

Interview Protocols

We designed several interview protocols to capture information about district assessment practices. A phone interview protocol was designed to identify the appropriate people to interview, the documents to collect, and the general procedures for hiring and supporting beginning teachers prior to the site visit. We also designed three interview protocols to capture information about assessments at each of the three major assessment points (i.e., hiring, continued employment and tenure, and support decisions). Each protocol included questions that focused on the processes, criteria and standards used to evaluate beginning teachers at that particular assessment point. In addition, other questions gathered information about the assessment's technical quality (e.g., training for assessors, use of standardized protocols, etc.), possible differential impacts, constraints, time required to conduct it, and use at other assessment points.

TABLE 3.2

LEA INTERVIEWS CONDUCTED BY TYPE OF COMMUNITY SERVED AND ROLE

	Urban	Suburban	Rural	TOTAL
School administrators	39	21	8	68
Mentor teachers	19	11	5	35
Other school staff	0	5	0	5
District personnel staff	12	3	0	15
District staff development staff	6	7	0	13
Other district staff	12	1	3	16
TOTAL	88	48	16	152

Each interview protocol was designed to take approximately one hour. In practice, however, the hiring and continued employment instruments each took slightly more than an hour to 1 1/2 hours for most respondents to complete, while the support protocol took approximately 45 to 60 minutes to complete. Actual interviews using these protocols varied in length depending on the knowledge and involvement of the respondent in assessment decisions, and whether there were other individuals who could also provide the information.[1] Sometimes more than one protocol was used in a given interview (e.g., hiring and support protocols).

---

[1] For example, in one very small district the superintendent possessed the most information about hiring, continued employment and tenure, and support policies and practices and answered questions on all three protocols. In other instances where an individual was knowledgeable about a very specific area such as district level support for beginning teachers, only a subset of questions on the support protocol were asked. Thus, interviews could range from roughly 45 minutes to as much as four hours, with most interviews about one-and-a-half to two hours in length.

Documents Collected

The collection of relevant documents concerning employment, continued employment and tenure, and support for beginning teachers was another major method used to obtain information on district assessment practices. Each district was asked to provide copies of official policies, procedures and other materials that would explain district hiring, continued employment and tenure, and support practices for beginning teachers. A list of the documents requested from each district is shown in Figure 3.1. For all districts in the sample, we collected application materials, as well as the procedures and forms used to evaluate teachers for the purposes of hiring and continued employment. In addition to these documents, other relevant materials were collected such as detailed handbooks and guides prepared for beginning teachers, district evaluations and studies of beginning teacher needs, and plans for initiatives related to evaluating beginning teachers such as new partnerships with universities or new standards for teacher retention.

Interviewers

Seven trained interviewers made visits to districts to conduct interviews and collect the relevant documents. For 7 of the 12 districts, only one interviewer conducted all the interviews in the district. In four districts with large numbers of interviews scheduled in a short period of time, two interviewers conducted the interviews on-site. For one very large district, four interviewers were employed to capture the range of settings within that district. All interviewers had previous teaching or administrative experience or considerable research and evaluation experience in California schools and districts. Interviewers were trained in two all-day training sessions in the use of the interview protocols and in the general strategy and procedures for conducting site visits.

Based on a review of the interviews conducted and documents obtained, the interviewer responsible for each site synthesized the data and completed a site summary questionnaire. During a debriefing session, interviewers were trained in completing these questionnaires and discussed common issues and patterns in assessment practices. Completed site summary forms were analyzed along with the documents obtained from each site. These analyses are the basis of the findings reported here.

FIGURE 3.1

LIST OF LEA DOCUMENTS

- Application forms (if different: elementary, secondary, bilingual)
- Job descriptions (elementary, secondary, other)
- Policy statements on official hiring procedures and forms for evaluation of applicants for employment
- Documents containing official policies or procedures governing the process of obtaining and/or screening applications
- Formal procedures or policies governing the face-to-face interview with applicants
- Written procedures for selection and training of staff who conduct the screening of, or face to face interview with, applicants
- Documents containing official policies, procedures and forms for evaluation of beginning teachers for continued employment
- Written procedures for selection and training of staff who conduct evaluations of beginning teachers for continued employment
- Official policies and procedures for the termination of nontenured teachers
- Forms for the termination of nontenured teachers
- Written procedures for selection and training of staff who evaluate teachers in the case of possible termination
- Documents containing official policies, procedures, and forms for the evaluation of teachers for tenure
- Written procedures for selection and training of people who evaluate teachers for tenure
- Policies and procedures for beginning teacher support
- Documents concerning staff development and other support for beginning teachers (e.g., handbooks or guides specifically designed for beginning teachers)

LEA Context

Features of the local setting establish the parameters for assessment by enabling or constraining district assessment practices. The manner and extent to which applicants are assessed during decisions about hiring, continued employment, tenure and support depend in part on several factors that we found affected the hiring process. For example, only applicants who have met state requirements such as passage of the CBEST can be hired. Similarly, collective bargaining agreements affect the numbers of new teachers that can be employed (e.g., seniority rights in many districts with declining enrollments can place a cap on the number of teachers that can be hired).

Three other factors that may influence the process of evaluating teachers for employment include: (1) a district's ability to attract and retain teachers, (2) rapid growth, and (3) the level of resources available for assessment. First and foremost, districts in our sample varied considerably in their ability to attract and retain teachers. While there is some evidence that a few districts could improve their recruitment efforts, for many of the districts in our sample this was a factor largely beyond their control. Contexts simply differ in their attractiveness to the majority of those entering the profession. Very remote and isolated districts, districts where safety and violence in the community are concerns, or districts where air quality or other aspects of the environment are unhealthy find it more difficult to attract high quality, or even credentialed, teachers. Such characteristics of the local setting influence the number and quality of applicants who apply and place obvious limits on the degree to which districts can be selective in their hiring practices.

Rapid growth in enrollments can also affect assessment practices by limiting the amount of attention districts are able to devote to teacher assessment due to the sheer volume of evaluations that must be conducted. Districts with limited resources that need to hire larger numbers of teachers may not be able to evaluate teachers as rigorously as they would like under other circumstances. At the same time we found that recent rapid growth which results in a staff relatively low on the salary schedule can allow more resources for assessment.

These factors in turn have an effect on the size and quality of the available applicant pool and the degree to which districts have the capacity to evaluate beginning teachers. They also determine the breadth and character of recruitment strategies. For example, we found when districts were able to be selective (that is, the number of teachers needed was small relative to the size and quality of the applicant pool), many districts recruited from their substitute, intern and/or student teaching pools -- sources that provided them with an

opportunity to evaluate applicants on the basis of their actual teaching performance in the district. Districts unable to be as selective but with reasonable levels of resources tended to rely on more active recruitment efforts out-of-state and overseas, while districts with fewer resources and more severe problems tended to have the highest number of staff teaching on the basis of an emergency certificate.

Finally, we also found regional or contextual differences in the characteristics districts look for in beginning teachers. For example, an applicant's potential to stay in the district, and his/her flexibility and ability to be a generalist instead of specialist with regard to subject matter, are characteristics more important to rural districts than to other districts. In districts with an aging staff, beginning teachers were frequently looked upon as a source of new ideas, energy and a vehicle for promoting change, and were selected accordingly.

Overview of Major Assessment Points

In this section, the processes and criteria used to evaluate prospective and beginning teachers at each of the major assessment points -- hiring, retention and tenure, and support -- are described. For each assessment point, the discussion focuses on: (1) the range of methods used, (2) the purposes and uses of various assessments, and (3) issues of technical quality.

Hiring Teachers

A review of the data on the hiring practices of districts in our sample suggests that the process of evaluating teachers for employment is quite complex and varied. We asked principals, district personnel staff, and others involved in hiring decisions questions about the specific procedures and assessments used in the hiring process. Our analysis indicates that three components or stages are common to the hiring process of all sample districts: (1) a procedure for initially screening applicants to determine who is qualified, (2) a process for selecting a pool of qualified applicants for a specific position, and (3) a final selection process to determine who to hire for a given position.

Within each stage, district assessment practices differed in the degree of complexity and the number of screens. For example, initial screening procedures could be relatively simple with district staff reviewing documentation to ensure that minimal state requirements had been met. Or the process could be fairly complex with the district initially screening applicants against both minimal state and district requirements, followed

by a series of additional screens such as group interviews coupled with observations to develop a qualified pool of applicants at the district level. We found most districts initially screen applications for completeness, and to see whether state requirements have been met, such as passage of the CBEST and a credential or major/minor in the subject area of the position for which applicants are applying.

Moreover, within the same method of assessment such as an interview, district interview practices differed in the format, length and use of standard questions. On the one end are districts where initial screening interviews are brief (10-20 minutes), fairly informal, and constitute a quick appraisal of an applicant's potential. Applicants are typically rated as outstanding, acceptable, or unacceptable, and this information may or may not be used in later selection processes. On the other end are districts with panels or teams comprised of district and school staff who use a standard set of interview questions to select qualified applicants following a thorough review of an applicant's written materials.

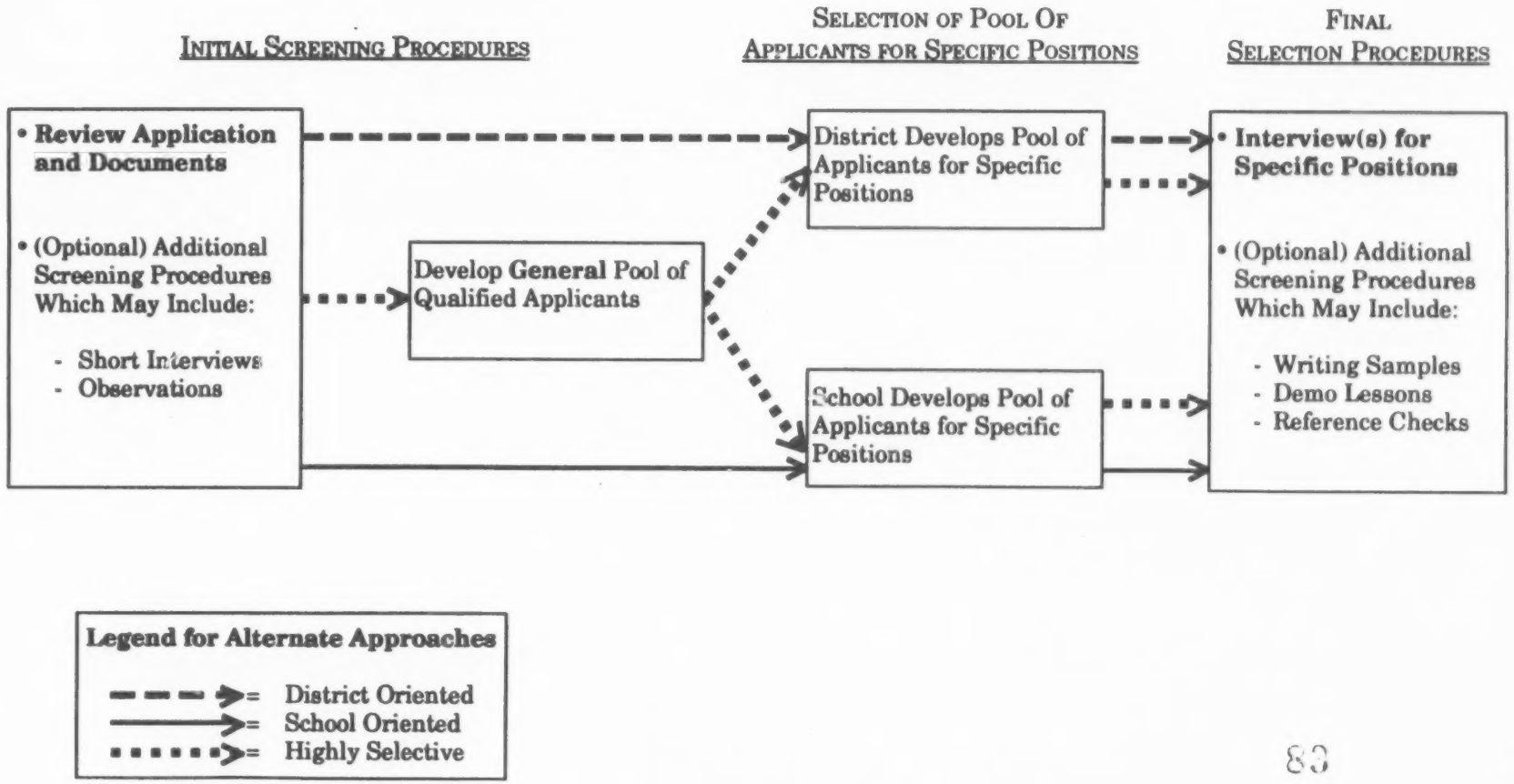
In addition, we found three major ways in which a teacher is hired for a specific position as well as a few minor variations. The general model of the hiring process and alternate approaches to evaluating teachers for employment in districts in our sample is depicted in Figure 3.2.

In general, district hiring practices differed along two major dimensions which distinguish the alternate paths to hiring teachers: the degree to which districts had the capacity to be selective about the teachers they hired, and the degree of school autonomy within the processes of developing a pool of applicants for a specific position and determining who to hire.

Capacity for selectivity. Our analysis suggests the degree to which districts have the capacity to be selective about who they hire is one distinguishing characteristic of district hiring practices that represents one of the three alternate paths to hiring teachers. Selectivity refers to the size and quality of the applicant pool relative to the number of teachers needed, while capacity refers to whether the district has the resources to evaluate the available pool of applicants.

In general, we found that districts who had the capacity to be more selective, typically developed a smaller, general pool of qualified applicants from which applicants were selected for interview at either the district or school level. Four districts in our sample screened applicants to develop a general pool of qualified applicants at the district level. In these districts, applicants were evaluated through a review of documents as well as through

FIGURE 3.2
Model of the District Hiring Process: Three Major Approaches


INITIAL SCREENING PROCEDURES
SELECTION OF POOL OF APPLICANTS FOR SPECIFIC POSITIONS
FINAL SELECTION PROCEDURES
• Review Application and Documents
• (Optional) Additional Screening Procedures Which May Include:
- Short Interviews
- Observations
Develop General Pool of Qualified Applicants
District Develops Pool of Applicants for Specific Positions
School Develops Pool of Applicants for Specific Positions
• Interview(s) for Specific Positions
• (Optional) Additional Screening Procedures Which May Include:
- Writing Samples
- Demo Lessons
- Reference Checks
Legend for Alternate Approaches
= District Oriented
= School Oriented
= Highly Selective

a series of additional screens, most typically through interviews conducted by a group of, rather than individual, staff.

In contrast, following brief initial screening procedures, the remaining eight districts moved directly to evaluating applicants to develop a pool or list of applicants for specific positions. In four districts, the development of a pool of applicants for specific positions occurred at the district level, while in the other four districts it occurred at the school level. In general, a second review of documents was the major screening device.

School autonomy. The degree of school autonomy in the hiring process, especially in the selection of a pool of applicants for specific positions, is another characteristic that distinguishes district hiring practices and represents the other approaches to hiring teachers shown in Figure 3.2. Districts in our sample were evenly split between those districts where school personnel constructed a list of applicants, and those where the district did. In **school-oriented** processes, schools typically developed a list of applicants to interview for a specific position by evaluating the applications of qualified applicants using predetermined criteria. In more **district-oriented** processes, several qualified applicants were usually selected at the district level and sent to schools for interviews.

Apart from these two major dimensions, no clear pattern emerged among districts in our sample, i.e., no approach to hiring was predominate among sample districts. Each of the three routes had four districts. Within the highly selective route, districts also differed in the degree of school autonomy in selecting a pool of applicants for a specific position. At this point, two districts had more school-oriented processes and the other two districts had more district-oriented processes.

However, from the limited evidence available it appears there may be elementary-secondary differences in the degree to which teachers participate in the hiring process. Through interviews with principals and teachers, we found that secondary schools often involve department chairpersons in the process of selecting and interviewing applicants because they are able to assess an applicant's knowledge of the subject matter as well as the applicant's fit with existing department strengths and weaknesses. In contrast, teachers at the elementary level usually did not participate in the hiring process.

Again, though the evidence is limited, interviews with principals suggest that clear district or school visions of the desired curriculum, preferred instructional strategies, or particular directions schools hoped to take, sometimes guide selection of teachers. For example, in one school we visited, the principal hoped to work on the science curriculum in

the upcoming school year. In hiring new teachers, the expertise a teacher could contribute in this area, and how the teacher's skills fit with those of other staff at that grade level, were important factors in his hiring decisions.

Range of assessment methods used. Across the 12 districts several methods of assessment are used to evaluate applicants for employment. Table 3.3 shows the range of methods used in employment assessments on a **district-wide** basis by the districts in our sample. Because schools have considerable latitude in the hiring process, some of these methods may be used on an idiosyncratic basis by principals in a given district. For example, several principals in district 10 mentioned observing an applicant teaching as one method of assessment, but observation was not a standard practice within the district, or a method used to assess all applicants. In contrast, district-wide methods are methods that are used systematically within the district, and typically are part of official district policy.

As shown in Table 3.3, a **review of documents** and materials submitted with a completed application, sometimes referred to as a "paper screening" of applicants, is a method used on a district-wide basis by all sample districts in hiring teachers. Our analysis of the information requested on district application forms used by the 12 districts reveals that at a minimum, all districts assess the following information from applicants: (1) educational background (including transcripts), (2) professional experience (including student teaching), (3) attestations about their character and/or work, (4) credentials held (or applied for), (5) positions sought with grade-level preferences, and (6) placement files. Districts differ, however, in how thoroughly these materials are reviewed, and in the assessment criteria and standards they use.

Interviews are the other typical method of assessing teachers for hiring used by all districts and schools in our sample. As part of the final selection procedure, all districts interview applicants for specific positions at least once, with interviews ranging from 20 minutes to 60 minutes in length, and averaging 30-40 minutes. Eight of the 12 districts in our sample also conduct a brief, initial screening interview lasting anywhere from 5 to 30 minutes as part of the initial screening process.

Our analysis of the data available indicates the use of standard forms for rating the applicant after the interview is a fairly common practice among districts, while the use of a standard list of interview questions is less common. Ten of the 12 districts in our sample routinely rate an applicant following an initial screening interview. Typically, applicants are rated along a 3- or 4-point scale such as outstanding, acceptable and unacceptable (though a few districts use 10-point scales). Eight districts (including those that use a list

TABLE 3.3

HIRING ASSESSMENTS:
METHODS OF ASSESSMENTS USED DISTRICT-WIDE

Methods of Assessment	Districts											
	1	2	3	4	5	6	7	8	9	10	11	12
Document Review	X	X	X	X	X	X	X	X	X	X	X	X
Interviews	X	X	X	X	X	X	X	X	X	X	X	X
• Using Standard Questions	X		X			X			X		X	
Reference Checks						X	X					
Writing Samples		X			X	X			X	X		
Demo Lessons						X						
Observation of Teaching	X											
School Tours				X				X				
Self-Evaluation Survey			X									
Lesson Plan								X				

of standard interview questions), also rate applicants following the final hiring interview. However, only 5 of 12 districts employed a standard list of questions for hiring interviews ranging from a list of 7 questions to as many as 30. In the other districts, the set of questions was left up to the individuals conducting the interview.

Although the method of **checking** an applicant's **references** is only used on a district-wide basis in two districts, the limited data we have available suggests that, in almost all districts, some principals check the references of the two or three top applicants for a position. Several principals reported that decisions about which applicant to hire among the top two or three applicants could hinge on these reference checks.

Some districts reported evaluating an applicant's ability to write legibly without making too many grammatical errors. But reviewing **writing samples** is a standard district-wide practice for only 5 of the 12 districts in our sample. Four of the five districts require applicants to provide a handwritten response to a standard prompt or set of prompts. In only one district were applicants required to provide an on-the-spot writing sample. Moreover, only one of the five districts had pre-established criteria for evaluating the writing samples.

As indicated in Table 3.3, districts employed several other assessment methods, but few of these methods were district-wide practices. For example, principals in several districts reported **observing applicants actually teaching** as the best means of evaluating an applicant's abilities and skills, though they neither had the time nor opportunity to do this consistently. This appears to be one of the main reasons why districts often prefer to recruit applicants from their substitute, student teacher, or teaching assistant pools. In only one district were selected applicants routinely observed using a standard observation form. In another district, a **demonstration lesson** is required of all applicants who are interviewed for specific positions.

Interviews also revealed two fairly creative methods of assessments used routinely in one or two districts. In two districts, applicants are routinely taken on **school tours** to judge the applicant's reactions to the students, the context for teaching, and/or to gain more information about an applicant (e.g., the applicant is "interviewed" along the tour). In one district, an applicant's spouse is routinely taken on a tour of the school and community as a method of evaluating the applicant's potential to stay in the district. Finally, one district requires applicants to complete a **self evaluation survey** as part of the application process in which applicants are asked to rate seven areas considered important to teaching and to justify or explain their ratings.

Purposes and uses of assessments. A thorough reading of a prospective teacher's materials can enable districts to evaluate applicants on several important constructs. For example, all districts routinely look for evidence of an applicant's classroom management and skills in disciplining students in student teaching evaluations, letters of reference, or their written responses to questions about educational experience. All districts also review transcripts for some indication of an applicant's subject matter competence and general academic abilities.

But our review of documents and interviews with principals suggest the criteria and standards used to evaluate most of these constructs are either not clear, or, if clear, tend to be minimal. For instance, individuals interviewed in several districts reported looking for evidence of an applicant's classroom management skills in terms of a strong recommendation from the university supervisor, and especially the master teacher, about classroom management. "If it isn't mentioned, it isn't a strength," was often the standard used. All districts also evaluate subject matter competence in terms of completion of required courses or an undergraduate degree or major in specific subjects, although a few districts go beyond that. When reviewing transcripts to assess general academic abilities, standards tend to vary. Some assessors simply look at grade point averages (GPAs) while others examine the perceived level of difficulty of the courses taken or trends toward improvement over time.

The ability and time to evaluate written application materials thoroughly varies tremendously from district to district, and school to school. In some instances, the information that can be derived from these materials is straightforward, for example, evidence of language fluency. But other characteristics such as potential to work with diverse student populations is more difficult to evaluate. Also, individuals appear to vary in their ability to glean information from the same materials. We found some individuals we interviewed more experienced and adept at extracting information about applicants that, from our perspective, looked marginally relevant. For example, one rural principal used information about the types of clubs and extra-curricular activities an applicant participated in during their undergraduate and teacher education preparation to draw inferences about the extent to which a person was a generalist or specialist in their subject specialty. When combined with information about an applicant's other hobbies and interests (e.g., fishing versus theater), and location and type of previous position, it also gave him some idea of the applicant's potential to stay in the district. This example also illustrates that different districts make different types of inferences from the same information depending upon their context and needs.

In general, whether or not districts employed a standard list of questions, most interviews appear to be aimed at assessing general pedagogical knowledge and skills, or personal attributes. From interviews and a review of materials used when interviewing applicants, it appears to be fairly standard practice to ask applicants questions related to general pedagogy, such as the ways in which they would handle certain student behavior situations, or to describe how they would organize their class for a particular subject. Personal qualities considered important are also assessed during an interview, though more frequently by subjective judgments rather than direct questioning. Many principals and others involved in interviewing applicants at the school level reported assessing an applicant's enthusiasm, maturity, and ability to communicate and relate well to others during the interview. More often the rule than exception, staff involved in district-level interviews also rate these personal attributes.

When principals and other evaluators were asked questions about the qualities or characteristics they look for in an applicant's interview responses or in screening applicants, most often they mentioned personal attributes. Characteristics such as enthusiasm, ability to relate to students, an understanding of the characteristics of a particular community, flexibility, positive outlook and maturity, and commitment to the profession were mentioned far more frequently than general pedagogical knowledge or skills. From the data available, this was especially true for districts which had difficulty staffing open positions.

Technical quality. With the exception of one district, we found very little training available for individuals who conducted evaluations, and what was available varied widely. For most assessment methods generally, and the two methods used on a district-wide basis in all districts specifically, little training was available. For example, with the exception of guidelines for identifying whether state or district requirements such as passage of CBEST or a minimum GPA have been met, in all but one district we found little evidence that individuals who screened written materials received any training or other guidelines on reviewing documents.

In general, with regard to interviews, in districts where hiring interviews were conducted by groups (either at the district or school level), criteria for assessing applicants tended to be more clearly defined. We found that individual principals conducting the school-level interviews are often neither trained nor required to use a standard list of questions or criteria; and even when required, their own criteria sometimes take precedence.

It should be noted, however, that the degree of rigor in the hiring process depends in part on the volume of teachers evaluated as well as the ability to be selective about who one

hires. For example, even though rural districts may be similarly isolated, a district that needs to hire only one or two new teachers each year can devote more time to evaluating applicants without the fear that a classroom won't be staffed in September as compared to another rural district which needs to hire several new teachers.

Continued Employment: Retention and Tenure of Beginning Teachers

Full-time certificated elementary and secondary teachers in California's K-12 public school system have certain continued employment and job security rights, but are on probationary status for their first two years. Teachers can be dismissed during the probationary period for reasons that need not (and in practice usually are not) disclosed to the beginning teacher. If no action is taken to remove the teacher prior to the end of the two years, the teacher becomes a permanent (i.e., tenured) teacher in the district. After gaining permanent status in the district, a teacher may be dismissed only for cause.

Though the frequency of evaluation varies from district to district, we found the basic process of evaluating beginning teachers for continued employment is fairly similar across districts. Beginning teacher evaluations can take two forms: formal and informal. Formal evaluations (which can include one or more assessment points) are those evaluations: (1) for which there is written documentation, (2) which are required by the district, and (3) which become part of a teacher's file. All other evaluations are informal. In accordance with the Stull Act, all districts in our sample formally evaluate a new teacher at least once annually. Two formal evaluations per year (usually one per semester) is the typical pattern among the districts in our sample (i.e., for 8 of 12 districts). Usually these formal evaluations contain areas of commendation (strength) and recommendations for areas of improvement. In addition, ratings of other job qualities such as attendance and participation in adjunct duties are also part of most forms. Only one district does not use a formal evaluation form that is shared with the teacher. In this district the formal evaluation consists of a letter to the teacher, which includes the recommendation for rehire or dismissal.

In general, these evaluations are based primarily on a number of formal and informal observations conducted by the principal, or, in large secondary schools, by assistant principals. Pre-observation conferences are usually held where, at a minimum, the forms to be used are reviewed and the processes of observation and evaluation are described. An observation is conducted and a post-observation conference is then held to discuss what was observed and/or the formal evaluation made. Details on the range of methods used to evaluate beginning teachers during hiring and the probationary period are discussed in the

following section. We briefly report here a few of the major findings related to the overall assessment practices of school districts for beginning teachers.

First, it is important to note that once hired, most school administrators generally make very little distinction in evaluations and assessment practices for beginning teachers regardless of experience and education. For example, in our interviews we found that most principals could not remember the backgrounds of particular teachers they evaluated without checking their records. Moreover, even with continual reminders during the interview, respondents still tended to mesh the problems and needs of all teachers new to the district. With the exception of support for interns in a few districts, support providers similarly could not differentiate among beginning teachers in their diagnosis or provision of support. As a result, it is impossible to cleanly separate our findings on assessment practices for beginning probationary teachers from other types of teachers new to the district.

Second, all districts in our sample used the same forms and procedures for evaluating and recommending beginning teachers for continued employment and tenure. Interviews revealed that while many principals may alter their standards of what they expect of beginning teachers in their first as compared to their second year (when tenure is awarded), these standards and the criteria emphasized vary considerably from principal to principal. For example, one principal's expectation of the degree to which a beginning teacher has clear goals for what students are learning and methods for monitoring that learning may differ for first and second year evaluations. But another principal might not use a differential standard between years or emphasize these criteria.

While most principals reported that evaluations are simply conducted for rehiring decisions and not differentiated for tenure, 2 of the 12 districts did attempt to make some distinction between tenure and continued employment. One district developed (and subsequently adopted) plans for evaluating teachers for tenure which set out evaluation criteria and standards against which to assess beginning teachers. Another district felt it was important to acknowledge the tenure decision as signifying entry into the profession. Though the evaluation criteria for the first and second year did not differ, a special reception is held for beginning teachers in the district who are recommended for continued employment in the second year (i.e., tenure).

Third, though the probationary period for beginning teachers is technically two years, in practice the period during which beginning teachers can actually be evaluated is shorter than two years. Teachers must be given preliminary notice by March 15 if their

contract is not renewed for the following fall semester. Therefore, the period in which to observe and evaluate beginning teachers is often only a year and a half. In addition, depending on when a teacher is hired, this period could be as short as 13 or 14 months. If a developmental perspective of beginning teacher professional growth is assumed, then the current time period in which to assess teacher knowledge, skills and abilities would seem to be inadequate.

Finally, though further exploration is necessary, the available evidence also suggests that there is very little difference in the procedures and criteria for evaluating beginning versus more experienced teachers. While the assessment practices for tenured teachers are outside the scope of this study, when cast against the growing consensus about the developmental nature of teachers' professional knowledge and skills, the need to differentiate assessment practices, or to extend the period in which beginning teachers are evaluated, warrants some attention.

Range of methods used. In Table 3.4, the major methods of assessment used for continued employment evaluations on a district-wide basis in districts in our sample are shown.

As shown in Table 3.4, in all sample districts the main method of assessment upon which these formal evaluations are based is **multiple observations**, including formal and informal classroom observations. The number of formal observations conducted ranged from one to six formal observations per year. Within our sample, two or three observations for each formal evaluation is the most common pattern (7 of 12 districts), with four districts requiring a single observation for each evaluation and one district requiring six observations for its annual evaluation. The length of an observation ranges from 30 to 60 minutes, and at least two districts script these observations and provide a copy for the teacher to review.

We also found that informal observations play a role that is at least as important, if not more so, than the formal observations. Staff and principals in all sample districts reported that informal "walk-throughs" of classrooms is standard procedure in their district, with the frequency and duration of these walk-throughs greatest at the beginning of the year to quickly identify any serious problems. In addition, many principals reported conducting informal observations outside of the classroom: on playgrounds, at lunch duty, in the library, during inservices, in grade-level or departmental meetings, and interacting

TABLE 3.4

CONTINUED EMPLOYMENT ASSESSMENTS:
METHODS OF ASSESSMENTS USED DISTRICT-WIDE

Methods of Assessment	Districts											
	1	2	3	4	5	6	7	8	9	10	11	12
Multiple Observations (Formal and Informal)	X	X	X	X	X	X	X	X	X	X	X	X
Review Lesson Plans	X	X				X			X	X		
Consult with Others						X			X	X		

with colleagues, students and parents. From the principal's perspective, these informal observations serve several purposes:

- to identify critical problems early in the year,
- to make teachers feel more comfortable with the observation process,
- to provide principals with information about teachers in varied contexts and their ability to handle the diverse responsibilities and activities that comprise a teacher's job, and
- to assure teachers that support and assistance is available.

The frequency and duration of these informal observations vary as much between principals in the same district as across districts. Only 3 of the 12 districts required

principals (or other evaluators) to spend a minimum number of hours observing, ranging from 5 hours per week for all classrooms to 20 hours for each beginning teacher per semester. In the remaining districts, how often principals conducted walk-throughs varied according to the amount of time available to the principal and the priority s/he placed on it. Of note, we consistently found that mentor teacher perceptions of the amount of time principals spend observing were generally lower than principal reports of the amount of time they spend observing. While the data available is too limited to draw firm conclusions, we suspect some of these differences may lie in how teachers and principals define observation and evaluation. Mentor teachers are more likely to focus on assessments made during observation in classrooms, while principals include a broader range of behaviors in and outside of the classroom as part of their observations.

Peer coaching, team teaching, staff meetings and other regular interactions among teachers also provide principals and staff with opportunities to evaluate a beginning teacher's skills, abilities and work styles, e.g., a teacher's comfort working as a member of a team. Principals reported these were key opportunities to evaluate important job-related skills that went beyond the classroom. For example, "fit" with other staff (ability to work with other staff as well as complement existing staff strengths) could be evaluated from observations of staff meetings and other interactions with staff. Observations of parent-teacher conferences provided principals with some insight about a beginning teacher's ability to interact with parents and relate to community needs.

Another method of assessment for continued employment used district-wide is a **review of lesson plans**. For five districts in our sample, a review of lesson plans prior to the scheduled observation is part of the formal evaluation process. However, within these districts, principals varied in the frequency with which they review plans, the standard to which they hold beginning teachers, and the use of information from these plans. For example, in at least two of these districts, lesson plans are also reviewed on a regular basis throughout the year. In addition, there is considerable variance among principals about the criteria used to judge these lesson plans. Some principals review lesson plans to see whether there are lesson objectives and, if so, the clarity of those objectives; other principals examine lesson plans to see how well the district curriculum is being met. Principals also varied considerably in their judgment about the degree to which lesson plans provided useful information for professional growth or evidence of expertise.

Consulting others for information is a third major method of assessment used by districts. Although consulting with others is a method used on a district-wide basis in only three districts in our sample, some principals in all sample districts reported routinely

soliciting information from as many people as possible. Mentor teachers, department chairs and assistant principals were frequently mentioned as sources of information, and several principals felt parent and student feedback was especially valuable. Other principals reported accepting, but not soliciting, information from others.

Though principals varied considerably in the degree to which consultation with others informed their evaluations of particular beginning teachers, the majority of principals consulted others when they observed particular problems or had specific concerns about a beginning teacher. For example, district or other school staff such as the department chair or assistant principal might be called in to observe a beginning teacher who appeared to be having substantial difficulty. Similarly, a principal might discuss with a mentor the areas in which assistance was already being provided to a marginal beginning teacher.

In our sample of districts, other methods of assessment are used to evaluate beginning teachers, but not in any systematic fashion. For example, in 9 of the 12 districts, at least one principal reported using indicators of student performance as one method of evaluating beginning teachers for continued employment. These measures included: reviewing student report cards and teacher grade books, examining CTBS scores, and observing or reviewing samples of student work. The standards for these criteria are not clear, and unfortunately, the data is insufficient to explore in-depth the ways in which they are used. Finally, feedback from parents and students, number and type of student referrals, and teacher performance in adjunct duties were also mentioned as sources of evaluative information by many principals. However, the use of these evaluation criteria appears to be idiosyncratic.

Purposes and uses of assessments. During the probationary period, multiple observations serve as the foundation of evaluations for continued employment in the districts in our sample. While state laws and collective bargaining contracts may shape the structure and content of formal observations, our evidence shows that what districts look for during formal and informal classroom observations is roughly similar -- e.g., over many different observations, most principals report initially looking for skills in classroom management, discipline and appropriate work behavior (e.g., arriving on time), followed by gradual improvement in other general pedagogical skills.

In all districts in our sample, evaluations for continued employment combine formative and summative assessments. Principals in all districts reported that while they conduct the formal conferences, observations and evaluations that are summative in nature (i.e., assessments directly related to decisions about continued employment), their real focus

is on the formative-supportive functions of assessments. This is especially true during the first year. Principals look for areas in which beginning teachers might need assistance, and they expect problems with classroom management and discipline. Once an area of need is identified, principals may provide support themselves, or more likely, refer the beginning teacher to district resources, mentor teachers and other school staff who can provide assistance.

In general, continued employment evaluations are used to identify areas of need as well as to inform subsequent decisions about dismissal. Usually marginal and unacceptable performances by beginning teachers tend to hinge on problems with classroom management and discipline. What is less clear from the available data are the standards used to determine when a beginning teacher is potentially in trouble, and how principals shift back and forth between their formative and summative roles. Most principals report that when a teacher's performance is marginal, they try a variety of strategies to provide assistance and support. But over time, if little or no improvement occurs, they bring others into the process to observe (either formally or informally), and/or they consult with other individuals such as a department chairperson, mentor teachers, other school administrators, and district staff who have had opportunities to observe or work with the teacher in question. Documentation efforts increase the longer the problem continues. In all districts in our sample, relatively few beginning teachers are dismissed. Reasons for dismissal are not disclosed, which makes it difficult for these evaluations to be useful to teachers in addressing their weaknesses.

Though evaluations are designed to provide beginning teachers with information about their strengths and weaknesses, there is some evidence that the feedback provided is not communicated well. From the data available, it appears that when evaluators find an area of weakness, they frequently link it to a strength in discussing it with the teacher so that a vague or mixed message may be conveyed. Moreover, the weakness is often posed as a suggestion rather than a specific area which might need improvement. For example, a principal concerned about the number of students on task might say, "I really like the material you covered in your lesson on X. How about trying to include a few more questions or activities to involve students more next time?" instead of saying, "You need to improve the way you involve students in your lesson."

Analysis of the data available also suggests that information collected during the hiring phase is rarely used in decisions for continued employment. For example, some district staff report giving interview notes or their judgments of an applicant's strengths

and weaknesses to principals. But principals in those districts rarely report, and more often deny, that this information is used in evaluations for continued employment.

Technical quality. Training for the principals and other school administrators who conduct the evaluations for continued employment is fairly limited and tends to focus on the process, rather than the content and standards, of assessment. At a minimum, the district usually orients principals in the use of the standard evaluation forms, and the legal and technical issues related to dismissal procedures. Five districts we visited require new administrators to attend a district inservice, most typically on clinical supervision. In two districts, regular principal meetings are used as an occasional forum for discussion of evaluation issues. At the other extreme, in at least one district a few principals could only recall an inservice being held several years ago. In addition, although nearly all (11 of 12) districts used standard evaluation forms that provide broad evaluation criteria, there is considerable latitude for principals to define specific criteria and standards. In part, this may be a function of the diversity we observed among schools within many districts.

In small districts, we found that some shared understandings do develop based on frequent interaction among the few (three or four) administrators. Likewise, the degree to which districts or schools have a vision of the curriculum or preferred instructional techniques can also guide evaluations by providing criteria against which to evaluate teachers -- e.g., cooperative learning strategies, team teaching, or an emphasis on a particular subject area that year.

Support

All districts in the sample provide beginning teachers with some type of support, although the type, frequency and method differ markedly from district to district. Two components of the California New Teacher Project focus on the issue of support and provide better descriptions of the range of methods of support provided to beginning teachers as well as analyses of the features of support activities that appear most effective. In this study, we focus on the range of assessments for support, i.e., the range of processes and practices employed by support providers to determine the type of support needed by beginning teachers and to monitor teacher development and progress.

Traditionally, observations that are the basis of the summative evaluation for continued employment and tenure are also used for formative support purposes. Typically, areas of strength and weakness are written, and then discussed during a post-observation or annual evaluation conference. In general, the principals interviewed acknowledged the dual

role they played. Most principals felt they leaned more heavily toward the support function than the summative evaluation function, especially early in the year. While principals claim to be more support oriented and do provide some support themselves, they more often refer beginning teachers to other sources of support and assistance, such as mentor teachers, district staff development personnel or activities, university courses, and other teachers or resources in the school and district. If there is limited improvement by a teacher experiencing difficulties in an area such as classroom management, principals increase the frequency and intensity of assistance, and more varied strategies of support are prescribed. If there is still limited or no improvement, principals begin more active efforts to document the evaluations and support provided.

Almost all of the mentor teachers we interviewed did not see their role as including an evaluation component. By law, mentors cannot be involved in the summative evaluation of beginning teachers. It is not surprising then that none of the districts had a feedback loop between support and evaluations for continued employment; that is, mentor teachers rarely give principals information about the support they provide. For example, with the exception of informing principals that assistance had been provided in the area specified, mentors did not typically disclose their judgments of the teacher's progress or development in that area. In this study, however, we tried to make a distinction between formative and summative evaluations, and to convey to the mentors interviewed that formative assessments are frequently conducted in order to determine the kind of support to provide or to monitor teacher progress. In six districts, mentor teachers clearly attempted to separate even this formative evaluative role from support by only reluctantly observing in classrooms or observing only when requested to do so.

Range of methods used. As a result of the law and various perceptions about support roles, the range of methods used to assess beginning teachers for support on a district-wide basis is limited and varies only by who conducts the assessment. We asked mentors, principals and others who provide support some questions about the strategies they use to determine the kind of support needed. The methods used for support assessments are listed in Figure 3.3. Those methods in bold-face are methods used on a district-wide basis, while methods in regular type-face are examples of the kinds of methods frequently mentioned by support providers, but not used systematically within a given district.

FIGURE 3.3

METHODS USED FOR SUPPORT ASSESSMENTS

Methods used by persons who provide support and conduct continued employment assessments:

- **multiple observations**

Methods used by other support providers:

- **referrals**
- **teacher request or self report**
- Other methods include:
  - targeted questioning
  - past experience
  - classroom observation
  - lesson plans
  - survey/needs assessment

For principals and others who conduct formal continued employment evaluations, **formal and informal observations** during the probationary period serve the dual purpose of determining areas in which new teachers need support and whether or not a teacher is rehired. Thus, principals tend to use the same methods discussed under continued employment and tenure to assess the need for support.

Because mentor teachers do not perceive themselves as evaluators, there are few methods used on a district-wide basis. In all districts in the sample, mentors consistently relied on **referrals** from the principal and/or on **requests for assistance** from beginning teachers themselves to determine the type of support needed.

Other methods of assessment that mentor teachers use to determine support are often more subtle and indirect. For example, many mentors reported relying on what **past experience** has taught them about what beginning teachers need. Some mentors did observe beginning teachers, but always informally, and usually with the beginning teacher's permission and/or at their request. Though the mentors we interviewed rarely cited **observation** as a method they used to diagnose needs for support, we suspect there are more opportunities for informal formative assessments than mentors may articulate, especially in schools where teachers interact on a regular basis. For example, team teaching situations,

grade-level or staff meetings, or the planning and implementation of an integrated curriculum across grade levels could provide mentor teachers with opportunities to observe and evaluate areas in which beginning teachers might need assistance. Other methods of assessment used to diagnose the needs, and monitor the progress, of beginning teachers reported by the mentors we interviewed included: reviewing **lesson plans**, conducting **surveys**, or **assessing needs** on a regular basis at group meetings.

One method mentioned by several mentors we interviewed was guiding teachers towards requesting assistance in particular areas through **targeted questioning**. The mentor often begins with a general question directed more to a beginning teacher's emotional state than to their technical skills, such as, "How did it go today?" The mentor then follows up on the teacher's response by either offering specific suggestions or by continuing questioning both to obtain more information on critical variables and to inspire the beginning teacher to generate alternative approaches toward addressing (or, sometimes, recognizing) problems. One district's mentor training includes instruction on targeted questioning techniques.

Purposes and uses of assessments. The purpose of all these assessments is obviously to provide beginning teachers with support and assistance. Like evaluations for continued employment, evaluations for support are designed to provide beginning teachers with information about their strengths and weaknesses. Similarly, like principals, mentors expect to find problems with classroom management early in the first year. In general, mentors initially look for problems with general pedagogical skills and knowledge. One recurring theme we heard from mentor teachers as well as from others who provide support is that beginning teachers must be able to manage the classroom first if student learning is to occur. We also found that assessments for support were typically based on a developmental perspective of a teacher's professional growth. One mentor teacher summed up the views held by many principals and mentor teachers when she commented, "I'm not looking for monumental, big step professional growth in the early years", she said, "but potential, receptivity to support, and adaptability so that we can develop the other skills."

Our analysis suggests that regardless of who provides the support, principals and mentors often communicate information in ways that may make it difficult for beginning teachers to get an accurate picture of their strengths and weaknesses. For example, analysis of mentor teacher descriptions of how they provide support suggest mentors indirectly guide beginning teachers to "seeing" an area of need, or couching the perceived areas of weakness in terms of suggestions. Instead of directly informing beginning teachers of an area of weakness, many of the mentors we interviewed would address an area that

needed improvement by first praising a teacher or commenting on something they liked, followed by a suggestion. For example, one mentor explained that she might praise the way in which bulletin boards and other aspects of the classroom environment looked and then suggest that the teacher might also post a list of classroom routines for students on the bulletin board. This was the mentor's indirect way of addressing a perceived problem in classroom management. Evidence from the teacher substudy supports this finding since many teachers did not recall any weaknesses being discussed in continued employment evaluations, though they could recall suggestions. There are clearly good reasons for using this approach. The major ones reported were to avoid overwhelming the beginning teacher, and to ensure that a beginning teacher's strong needs for positive support were met.

Technical quality. Because principals conduct support assessments at the same time as summative evaluations for continued employment, the findings on training for evaluators of continued employment and tenure discussed earlier also apply here. From the limited data available on training for mentor teachers, it appears that while training for mentors is considerable, specific training on assessment strategies is less commonly reported by mentors. However, this does not mean that mentors do not receive such training. One problem in interpreting the data available is that because most mentor teachers do not perceive themselves as evaluators, they are less likely to report related staff development activities in this area. Another difficulty we encountered is that while mentor teachers were asked about the training they received for offering support to beginning teachers, it was difficult to discern in their responses whether, and to what extent, strategies for determining or diagnosing support were covered as part of the general training that mentor teachers receive. In at least two districts, mentor teachers receive the same inservice that principals do. In at least one other district, mentor teachers are trained in targeted questioning techniques that guide beginning teachers into identifying their own areas of need.

Focus of Assessments

In interviews with people who are involved in hiring decisions, we asked questions about the qualities and characteristics they look for in screening the potential applicant, in selecting applicants to interview, and in applicant responses to interview questions, probing for information on whether subject-matter knowledge, pedagogy, content pedagogy, and other skills were assessed. Similarly, in interviews with people who are involved in continued employment and support decisions, we asked questions about the knowledge, skills and abilities they looked for when evaluating a beginning teacher during the probationary period or to determine if s/he needed support. These questions, in conjunction

with a review of the policies, procedures and forms used to evaluate beginning teachers for hire, continued employment, and support are the primary basis for the findings presented here.

Subject-matter Competence

We found that subject-matter competence is usually evaluated at the point of hiring in very limited, often perfunctory ways. In the initial screening process, all districts rely on an applicant's university-based training and completion of courses as the major indicator of subject-matter competence. Other than a review of transcripts, subject-matter knowledge is evaluated in interviews with questions about content pedagogy (see section that follows). There is some evidence that evaluation of subject-matter preparation is more prevalent at the secondary than elementary level, in that secondary teachers, especially department chairpersons, are more often and more consistently involved than elementary teachers in hiring interviews. We found that in secondary schools, some principals often relied on these teachers to evaluate an applicant's breadth of knowledge in a particular curricular area and to assess how well an applicant's areas of expertise complemented those of existing staff. Not surprisingly, however, the criteria and standards used are highly variable among schools. Assessment of subject-matter knowledge (independent of content pedagogy) at the other points of retention and support is very limited. Overall, unless there are glaring deficiencies in a practicing teacher's subject-matter competence, it is rarely assessed in beginning teachers after they are hired.

Pedagogy

General pedagogical knowledge, skills and abilities are assessed to varying degrees at all three assessment points, but appear particularly critical in evaluations for continued employment. All districts in our sample make numerous efforts to assess teachers' general pedagogical skills and knowledge, through multiple methods of assessment when possible. While the technical quality of these assessments is not high, and standards are not always clearly defined (e.g., the "I'll know it when I see it" variety), there does appear to be some agreement on the very general criteria as evidenced by the broad categories typically listed or described on evaluation forms (e.g., appropriate instructional activities, progress of students toward goals, etc.) and by the responses to questions we asked about the important abilities and skills evaluators look for in assessing beginning teachers.

Knowledge and skills in classroom management, organization, and discipline are the main pedagogical constructs evaluated at the hiring, retention, and support stages in all

districts irrespective of the teaching context. All districts look for indirect evidence of these skills at the point of hiring. At the hiring stage, districts usually assess, and sometimes rate, applicants on general pedagogical knowledge and skills based on responses to interview questions about typical classroom or school situations. There appears to be a widespread consensus among evaluators for continued employment (school administrators) and other evaluators of support (mentor teachers) that without basic skills in classroom management and disciplining students, no learning occurs. Most also agree that given time, other pedagogical skills and curriculum content can be learned on the job. Principals and others who conduct continued employment evaluations, as well as support providers, expect to find problems in these areas and typically focus on these skills first.

Evaluation of other pedagogical constructs is more varied and less systematic within and across districts.

Content Pedagogy

In general, assessment of content pedagogy is fairly limited and not a major focus at any of the three major assessment points in our sample of districts. Our data suggest that some effort is made to assess content pedagogy in at least six districts, primarily through interview questions about how an applicant might teach a particular subject or unit. However, established criteria or standards by which to judge responses are rare, and therefore highly variable within and across districts. From the limited data available, we found that content pedagogy is more likely to be evaluated in districts with a strong district curriculum. In particular, districts tended to focus on content pedagogy if there was a particular subject area that the school or district was focusing on that year. Insofar as a district's curriculum is aligned with the state curriculum frameworks, teaching in congruence with the relevant curriculum framework is assessed during continued employment and support evaluations.

Other Constructs

Our analysis of interviews and materials suggest that all districts in our sample assess general academic skills. Apart from passage of the CBEST, districts review transcripts to assess an applicant's general academic abilities. However, criteria and standards vary considerably. Some individuals simply look at an applicant's GPA, while others examined the perceived level of difficulty of the courses taken or a pattern of improvement over time. Only one district used a minimum GPA as a screening device. Districts also evaluate writing skills at least minimally. Individuals in most districts

reported reviewing completed applications for legibility and correct usage of grammar. But in only five districts were applicants required to respond to a standard set of prompts in their own handwriting, so that responses could be assessed systematically. In only one of these five districts were criteria by which to judge applicant responses available.

We also found that most districts evaluate applicants on personal attributes that provide indicators of an applicant's **potential** to become a competent teacher. During hiring evaluations, most districts were interested in assessing personal characteristics that indicated their **potential for the profession** and their likelihood of remaining in the profession. Thus, along with general academic abilities, characteristics such as liking children, enthusiasm, maturity, good interpersonal and communication skills, commitment to teaching, etc. were frequently mentioned as important criteria. After a teacher is hired, most districts also assessed a teacher's **potential to improve** and looked for personal qualities, such as willingness to learn, receptivity to support, openness to suggestion, and flexibility, that would enable a teacher to develop pedagogical and other skills.

Though such indicators alone may first appear inconsequential, they do support a developmental perspective of a teacher's professional growth. In addition, these characteristics in combination were used to assess more complex constructs such as the potential to teach within a specific context. Characteristics such as living in the community, commitment to the achievement of all children, ability to work with children, ability to relate to parents, willingness to be involved in community activities, and awareness and acceptance of cultural and ethnic diversity together provided some districts with an indication of an applicant's likely success and longevity in their district.

Finally, a few other skills frequently assessed by districts included an applicant's fluency in another language, and expertise and potential interest in extra-curricular activities. All districts in our sample appear to obtain information about an applicant's fluency in another language early in the hiring process. But only five districts request this information on application forms, and only two districts use standard interview questions that address language and bilingual issues. Unfortunately, our data is not clear about whether and how language fluency is verified or evaluated. Similarly, in all districts in the sample, some schools, especially secondary schools, assess an applicant's expertise and potential interest in coaching or sponsoring student organizations and activities based on information gathered about extra-curricular activities. But only six districts assess these skills on a district-wide basis.

Differential Impact of Assessment Practices

In our interviews with individuals involved in hiring and continued employment decisions, we asked respondents whether they gave any consideration to underrepresented groups (such as specific gender or ethnic groups) in evaluating beginning teachers for hire or continued employment, and to describe how the procedures or standards for evaluation differed.

An analysis of this data suggest that at the hiring stage, some special consideration is given to hiring beginning teachers from underrepresented groups. Ten of the 12 districts in our sample reported giving special consideration to underrepresented groups through more active recruitment efforts and selection procedures. But most districts also reported that hiring criteria and standards rarely differ. In general, districts gave special consideration to applicants from groups that would alter the existing staff ratios to better reflect the student population. Most often districts were looking for bilingual, especially Spanish-speaking, applicants, though in at least one district, Anglo applicants were sought. Other factors being equal, individuals in several districts reported male, particularly minority, elementary applicants were actively sought while one district looked for female applicants for secondary positions.

In evaluations for continued employment, however, special consideration is rarely given to beginning teachers from underrepresented groups. In only two districts were there reports of different standards, and in both cases, principals modified their standard on oral English skills because they felt the teachers' cultural background and fluency in other languages outweighed their English deficiency.

Costs of Assessments

This section addresses the issue of the cost estimates of conducting assessments for hiring, continued employment and tenure, and support. Calculating these costs is extremely difficult because of the complexity of the processes of evaluating beginning teachers, especially for hiring assessments where evaluations are often conducted at both the district and school level. We tried to collect comparable data about the costs of conducting assessments. However, this information is not compiled by any of the districts in our sample. Respondents found it difficult to provide any cost estimates, except in terms of time spent. Some respondents reported time spent per applicant, while others reported time spent per position advertised or in terms of full-time equivalent (FTE) employee time. In addition, the lack of consistency in how these assessments are conducted even within a

given school made it difficult for assessors to accurately estimate their time. For example, the amount of time devoted to assessments is affected by other demands on the assessor and potential assessors (e.g., teachers) which could vary from day to day. As a result, we were able to obtain some rough estimates from nearly all districts, but the data are not comparable across districts. Therefore, costs are reported in terms of the range of time spent conducting assessments in the simplest case and the most complex case at each of the three points. No modal cost estimate is reported because of the extreme variability in district assessment practices.

Hiring Assessment Costs

The model of the hiring process presented earlier illustrates the variation and complexity in district hiring procedures. We found two factors key to examining the issue of assessment costs for hiring: (1) the number of people conducting the assessment (one person versus a committee) and (2) the steps in the assessment screening process. Costs are described separately for initially screening applicants and for determining who to hire (including selecting who to interview, conducting the interview, and synthesizing information to decide who to hire).

First, the time involved in initially screening applicants depends on the number of people involved.[2] In the simplest case, one person simply screens all applications for completeness and to see that certain requirements are met. In the most complex case, two committees (each comprised of eight people) screen all applications to develop a general pool of qualified applicants. In both cases, additional clerical assistance is needed for computer entry of information, followup, establishing personnel files, etc. In our sample of districts, estimates ranged from about 15 minutes to two hours per applicant, depending on the number of steps in the screening process and people involved in the process.[3]

In addition, though not factored into this estimate, the number of applications evaluated relative to the number of people conducting the assessments affects the overall amount of time each district spends. In our sample, this ratio ranged from one person

---

[2] Actual monetary costs will also be affected by the role of the individual conducting the screening. We found this could range from a clerk to the personnel director or superintendent.

[3] Recruitment efforts could affect initial screening costs by increasing the number of applicants that need to be screened, and several districts in our sample spent a considerable amount of time recruiting out-of-state or even overseas. Data on recruitment efforts were outside the scope of this study.

evaluating 37 applications for two positions, to two people evaluating 3,500 applications for 200 positions.

The interview portion of the hiring process takes the greatest amount of time. A group interview procedure typically increases costs because additional time is needed for discussion and synthesis of opinions. The number of interviews also increases the amount of time spent. In the simplest case, typically a principal will interview two or three people sent by the district for a specific position. In the most complex case, two committees (one committee of four people and another committee of eight people) interview 15 applicants and jointly select applicants to hire. In most cases, additional time is spent in preparation and/or followup activities, such as selecting applicants to interview and checking the references of the top two or three applicants. In our sample of districts, estimates of the amount of time required to hire a teacher for a specific position ranged from 2 hours to 50 hours per position.

Continued Employment Assessment Costs

In continued employment assessments, formal evaluation costs are relatively simple to calculate. Typically, the time involved in conducting the formal evaluation includes a pre-conference, observation(s), write-up, and post-conference. In the simplest case, there is a single observation and evaluation (in the form of a letter informing the applicant of the retention decision) in a year. In the most complex case, two formal evaluations composed of three observations each are conducted annually. Estimates of the time spent on the formal evaluation process range from 1 1/2 hours to 20 hours per beginning teacher. One factor not taken into account in this estimate of the formal evaluation is the time associated with training assessors, or the additional time spent evaluating teachers who are experiencing difficulties.

The time involved in informal observations is more difficult to calculate. Two major factors affecting the time involved in informal observations are the number of beginning teachers to evaluate and the extent to which beginning teachers are experiencing difficulties. The number of beginning teachers to be evaluated in our sample of districts ranged from 1 to 14 beginning teachers per school. Most principals reported spending a considerable amount of time with marginal teachers and less time with teachers performing adequately or well. For these reasons, principals had considerable difficulty providing a general estimate of the time associated with informal evaluations, especially just for beginning teachers. The best estimates we were able to obtain for formal and informal assessments in our sample of districts ranged from 2% to 40% FTE for a principal.

Support Assessment Costs

From the available data, we found it impossible to estimate the time involved in assessing beginning teachers for support purposes. Most mentor teachers (who provided most of the support to beginning teachers in nearly all of our districts) do not perceive their role as having an evaluation component, even a formative evaluation component. Therefore, mentor teachers were not able to separate the time involved in assessing a teacher to determine the type of support to provide from the time they spent actually providing support.

Also, while mentor teachers were often required to spend a certain amount of time per quarter/semester providing support to teachers, in our sample of districts most mentor teachers still carried a regular, or occasionally slightly reduced, teaching load. In addition, many mentors believe that emotional support is necessary for beginning teachers and make themselves available by phone at all hours. As a result, mentor teachers had a difficult time estimating the percentage of time they spend supporting, let alone assessing, beginning teachers.

Contributions of LEA Assessments to Staff Development Decisions

Earlier discussions about the formative assessments made as part of the continued employment evaluations pointed to the use of information from these evaluations for determining the kinds of support to provide beginning teachers. We also gathered data on the contributions of hiring assessments to staff development decisions. In interviews about the hiring process, we asked respondents questions about whether information from the hiring process was ever used to identify areas of potential weakness for staff development and whether the teachers hired were given any feedback about the strengths and weaknesses that were identified during the hiring process. Similarly, within interviews about assessments for support of beginning teachers, we asked respondents questions about whether information from previous evaluations such as the NTE, student teaching evaluations, or information learned during the hiring process was used to identify areas in which to offer, or provide training for offering, support to beginning teachers.

From these interviews, we found that information obtained from the hiring process is used to guide the professional development of individual beginning teachers, and less frequently, to structure support and inservice activities in schools. In 10 of the 12 districts in our sample, principals or others responsible for conducting hiring interviews reported that they used information identified during the hiring phase to guide teachers into various

staff development options or support activities. This might be done at the point that the teacher is hired. For example, several principals reported using the information in conferences with beginning teachers to develop individual professional development plans at the start of the year. Principals also reported using information from the hiring process to plan school-level inservices or to structure support activities for beginning teachers.

Yet interestingly, specific information about a beginning teacher's strengths and weaknesses is less frequently provided to the beginning teacher. In response to the questions about providing feedback on strengths and weaknesses identified during the hiring phase, some principals and districts reported providing this information only when asked by the beginning teacher to provide it. More commonly, principals provided feedback on the successful applicant's strengths, not weaknesses. In cases where feedback on weaknesses was discussed, it was combined with a discussion of their strengths, or phrased in non-threatening ways. For example, in one district, principals routinely advise successful applicants of their strengths, and then point out areas where the applicants need additional work.

From the data available it also appears that information from assessments conducted in other settings, such as student teaching evaluations, is rarely used to inform staff development decisions. None of the people we interviewed reported using these sources of information to guide the professional development of individual beginning teachers, or to structure staff development or support activities. In some cases, this information was simply not available to principals or others involved in conducting hiring interviews at the school level.

Summary and Conclusions

The LEA substudy examined the range of ways beginning teachers are typically evaluated in hiring, continued employment and tenure, and support assessments in local public school districts in California. This section lists the major findings and summarizes the overall conclusions about LEA assessment practices.

Summary of Major Findings

Methods of assessment:

- Document review and interviews are the two main methods of assessment consistently used by all districts in hiring decisions.

- For continued employment, tenure, and support decisions, observation is the major method of assessment used on a district-wide basis by all districts.

- For support decisions made by support providers who do not conduct continued employment evaluations, referrals and teacher requests for assistance are the major methods used district-wide.

- Other forms of performance assessment such as demonstration lessons, writing samples, and lesson plans are used at various assessment points, but their use varies widely within and across districts.

Purpose and use of assessments:

- Information obtained from LEA-conducted hiring assessments are often used to guide the professional development of individual beginning teachers and to structure inservice and support activities; however, information from assessments conducted in other settings, such as NTE tests or student teaching evaluations, is rarely used to inform staff development decisions.

- In all districts, assessments of teacher performance and needs for continued employment serve both formative and summative purposes, though the standards used to determine when a teacher's performance shifts from marginal to unacceptable, and how assessors shift between formative and summative roles, are less clear.

- Generally, whether principals or mentors are the ones conducting formative assessments and providing support, it appears information is often communicated in ways that may make it difficult for beginning teachers to get an accurate picture of their strengths and weaknesses.

- Similarly, information about the strengths, and especially weaknesses, of beginning teachers identified during hiring assessments are either not communicated at all or are communicated in ways that are not useful for beginning teachers.

- Most assessors and providers of support have a developmental perspective on a teacher's professional growth. In general, however, given the limited time and resources available for supporting beginning teachers, formative assessments

are used most frequently to assist teachers performing marginally, with limited use for teachers performing adequately.

Technical quality and rigor of assessments:

- The extent and type of training for assessors who conduct evaluations for hiring and continued employment varies considerably, with most training tending to focus on the process, rather than the content and standards, of assessment.

- While nearly all districts use standard evaluation forms that provide broad evaluation criteria for continued employment assessments, and many districts use standard rating forms to summarize information from hiring interviews, there appears to be considerable latitude for principals to define specific criteria and standards.

- Looking across the assessments at various points, if there is rigor in assessments, generally, it is at the hiring stage for most districts.

- Most districts rely on multiple measures to assess a teacher's skills in classroom management.

- The degree of rigor in the hiring process depends in part on the volume of teachers evaluated as well as the ability to be selective in hiring.

Focus of assessment:

- Irrespective of the teaching context, all districts routinely evaluate knowledge of and skills in classroom management and discipline.

- Subject-matter competence is usually evaluated in very limited, often perfunctory, ways, primarily at the point of hiring.

- General pedagogical knowledge and skills, particularly in the area of behavior management, are assessed to varying degrees at all assessment points, but they are particularly critical to evaluations for continued employment.

- Generally, content pedagogy is assessed in limited ways; it is neither a major focus

of assessment nor consistently assessed within and across districts.

- Personal attributes that relate to a person's potential to develop and become a competent teacher are considered important constructs that are usually subjectively assessed by all districts.

- Other constructs are also assessed, but in uneven and limited ways. For example, while many districts attempt to assess knowledge of students, the assessments are generally low in technical quality and fairly unsophisticated, e.g., questions in hiring interviews focus on a person's familiarity and/or experience working with particular types of students.

Contextual factors affecting assessments:

- Features of the local setting that are largely beyond a district's control, such as the attractiveness of the context to the majority of those entering the profession, affect the number and quality of applicants who apply, and place limits on the degree to which districts can be selective.

- Rapid growth in enrollments can affect assessments by limiting the amount of attention districts are able to devote to teacher assessment.

- Limited resources affect the scope and rigor of assessment efforts, especially among districts which need to hire large numbers of teachers.

Conclusion

In conclusion, patterns in LEA assessment practices, i.e., consistent use of particular methods, criteria and standards, are difficult to discern given the level of school autonomy and considerable latitude allowed principals, mentors and others involved in beginning teacher assessments in defining specific criteria and standards. When combined with the variability in district needs and settings, one finds little consistency in LEA assessment practices. To a certain extent, this latitude and variability seem necessary to accommodate the diversity we observed among schools within many districts and between districts. At the same time, this lack of consistency means variable levels of expectations for beginning teachers both within and across districts in California.

Against this background, it is not surprising that the technical quality of these varied assessments is in most cases low. While there is considerable variation across districts, in general, training for those who conduct continued employment, and especially hiring and support assessments, is fairly limited. In particular, we found that criteria and standards were not clear or well defined. Even for skills deemed critical by all districts such as classroom management, where there is some agreement on general broad criteria, standards are not clearly defined and tend toward minimal performance expectations.

In addition, the usefulness of assessments to beginning teachers' professional growth is often limited. What time and resources are available tend to be invested in assisting teachers who are performing marginally, with limited support given to those teachers performing adequately. In addition, potential resources such as information from assessments conducted outside of the district (e.g., student teaching evaluations or NTE scores) are not always available or used by assessors at the school level.

Overall then, the LEA assessment practices currently in use for evaluating beginning teachers are not consistent, particularly rigorous, thorough, or a critical factor in beginning teachers' long-term professional development.

CHAPTER 4:

THE STATE ASSESSMENT PRACTICES SUBSTUDY

BEST COPY AVAILABLE

CHAPTER 4:

THE STATE ASSESSMENT PRACTICES SUBSTUDY

The assessments of prospective and beginning teachers conducted by IHEs and LEAs were described in the two previous chapters. The substudy described in this chapter concentrates on the State assessments administered by the Commission on Teacher Credentialing (CTC) for credentialing purposes. State credentialing requirements include demonstration of proficiency in basic skills, knowledge of the subject matter of the field(s) covered by the credential sought, and adequate professional preparation. Although the stage of a teacher's career at which corresponding assessments are administered varies, the common feature is that all teachers must have passed these assessments prior to receiving a multi-year California teaching credential.

This chapter examines the State assessment practices using the same categories of analysis which were applied to IHE and LEA assessment practices. The data reported, for the most part, are taken from published reports from the Commission on Teacher Credentialing or its contractors, test bulletins describing the assessment instruments, and other public documents. The chapter includes eight sections. In the first section, the sample and the methodology employed are briefly described. The second describes the broader context that shapes current State assessment practices. The third section provides a descriptive overview of the assessment practices which are used in credentialing teacher candidates, followed by a discussion of the focus of the State assessments. The fifth section discusses the use of the State assessment practices in IHE and LEA decisions. The sixth and seventh sections address the differential impact of assessment practices on teacher groups and costs respectively. The chapter concludes with a brief summary of conclusions about State assessment practices.

Description of Sample and Methods

Sample Description

Unlike other substudies, no sampling was required, since only one organization is involved at the State level in assessment practices. The requirement of uniform, standardized assessment practices at the State level also simplified the identification and analysis of existing assessment practices, as there is little variation in assessment practices

to capture. Indeed, great care is taken to ensure that for each teacher candidate the assessment practices are uniform and applied in a standard manner. The following discussion covers each assessment practice connected with issuing a multi-year teaching credential to regular and bilingual multiple subject (elementary) and single subject (secondary) teachers.

Methodology

Existing State assessment practices are well documented and several important aspects have been the focus of previous research. Therefore, this substudy differs from the others in that little original research was performed. Instead, this substudy relied on published descriptions and studies of the major assessment instruments, supplemented by interviews with CTC staff.

Initially, CTC staff identified major documents to be examined with respect to the major State assessments. After an initial examination of the documents, relevant CTC staff were contacted to answer questions about specific assessment procedures and/or instruments. Often, this brought additional documents to mind, and the process was repeated. In addition, CTC staff were interviewed about the legislative context in which specific assessment instruments or foci were mandated to further understand the purposes for which the existing assessment practices were developed.

An outline of topics to be addressed and frequent consultation with writers of other substudies helped ensure that the analysis of State assessment practices would parallel those conducted for the IHEs and LEAs.

State Context

The Commission on Teacher Credentialing (CTC) is charged with the development and administration of the regulations governing the issuance of teaching credentials in the State of California. Both statutory requirements and widespread consultation of teachers, administrators, and teacher educators inform the development of the credentialing requirements. The CTC reviews evidence submitted by teacher candidates as to their fitness for teaching, and determines whether or not to award a credential to each applicant. Although the credentialing requirements are designed to allow considerable variation in terms of individual teaching styles and philosophies, even within subject area, the CTC oversees the development and administration of several uniform statewide assessment instruments.

This section describes the history of the basic skills requirement, the subject-matter knowledge requirement, and the general and content pedagogical skills requirement implemented in the review of credential applications.

Basic Skills Proficiencies Requirement

In 1976, legislation passed which required students to show that they could read, write, and do simple math in order to receive a high school diploma. Shortly afterward, legislators began receiving reports from parents that their children's teachers lacked these same skills. Several districts reported that teacher applicants failed district-devised basic skills tests. The author of the student proficiency test legislation, then-Assemblyman Gary Hart, developed parallel legislation requiring prospective teachers to pass a proficiency test in the same skill areas at the collegiate level. The skill areas were chosen to parallel those required of students. The mathematics skill requirement inspired some debate among legislators as to whether or not it should be included. However, the movement in education was toward an integrated curriculum at the elementary level and an interdisciplinary curriculum at the secondary level. Therefore, teachers might be required to use mathematical skills in any subject, so the mathematical skills requirement was retained. The resulting legislation, California Education Code Section 44252, enacted in 1981, required that all prospective teachers demonstrate proficiency with respect to basic reading, writing, and mathematics skills. The Educational Testing Service (ETS) won the competitive bidding to develop such a test, known as the California Basic Educational Skills Test or the CBEST. The CBEST is not designed to measure teaching skills, but to insure that teachers working with school children can read, write, and compute at the basic levels required in college. The CBEST has been in use since 1982.

Subject-matter Knowledge Requirement

Prior to 1960, not only was there no subject-matter knowledge requirement, there was also no subject-specific credential. Instead, there were two regular teaching credentials: a general elementary credential and a general secondary credential. This credentialing system enabled districts to assign teachers to any class in any subject area within the appropriate grade level. With the launching of Sputnik and the attendant emphasis on math and science education, a concern with the extent of subject-matter knowledge of math and science teachers arose. The Fisher Act was subsequently passed which eliminated undergraduate majors in education from all California teacher preparation programs. Secondary teachers were only authorized to teach in the area of their major or minor, severely limiting the subjects they could teach. For instance, a history major with a minor

in psychology could teach history or psychology, but not government or sociology. This created tremendous staffing problems for districts.

In response to these problems, the 1970 Ryan Act modified the requirement to demonstrate subject-matter knowledge by redefining and broadening the focus on a "subject" to include "subsumed subjects" in a more general subject area. For example, the English subject area was redefined to encompass not only literature and composition, but also journalism, drama, and speech. The legislation favored standardized examinations as the preferred method of demonstrating subject-matter knowledge. The intent was to develop subject-matter examinations for prospective California teachers. However, in the interim before the development of such examinations, the legislation provided for satisfaction of the subject-matter knowledge requirement through existing examinations meeting specific criteria or through completion of an approved set of courses addressing topics specific to each subject area. After a review of existing subject-matter examinations, the set of National Teacher Examinations (NTE) was chosen to measure the subject area mastery of credential applicants. The adopted NTE tests adopted focussed on subject-matter knowledge, but also included some items on pedagogy and/or subject-specific pedagogy as well.

Multiple subjects credential. Since 1984, the NTE Test of General Knowledge, a subpart of the NTE Core Battery Examination, has been used to measure subject-matter knowledge for the multiple subjects credential, which covers a number of subjects contained in the elementary curriculum. This test consists of four sections: literature and fine arts, mathematics, science, and social studies. The test was designed to represent courses and subjects that contribute to a teacher's general education. It came the closest of any existing test examined to representing the basic knowledge at the liberal arts undergraduate level in core subjects that was required as preparation to teach multiple subjects at the elementary level. However, dissatisfaction with the adequacy of this test has grown over the years.

Several factors led to the recent solicitation of bids to develop a new examination specifically tailored to the California multiple subjects credential. Numerous Model Curriculum Guides and Curriculum Frameworks addressing instruction in the elementary grades had been issued by the California Department of Education. In response to the resulting substantial changes in the elementary school curriculum, the CTC adopted new standards for subject-matter preparation for the multiple subjects credential to be implemented no later than the Fall of 1991. The new standards require demonstration of breadth across the subjects in the elementary curriculum, depth in one subject area chosen as a concentration (area of particular expertise), and interdisciplinary knowledge needed to

integrate across the subjects in the elementary curriculum. At this point, it was apparent that substantial differences existed between the NTE Test of General Knowledge and the new subject-matter knowledge standards for the multiple subjects credential. The Commission decided to request proposals to develop a new examination that better reflected the new requirements.

The resulting examination, tentatively titled, "The Subject Matter Assessment of Elementary Teaching," was pilot tested in the fall of 1991, and is scheduled to be field tested in the spring of 1992. After completion of the field testing, the performance of the various modules will be analyzed, and the composition and the length of the examination will be determined.

Single subject credentials. Parallel changes in the secondary curriculum led to questions about the appropriateness of the NTE Specialty Area Tests as measures of subject-matter knowledge for the secondary subject areas. In 1987, the appropriateness of the fifteen NTE Specialty Area Tests used in California to satisfy the subject-matter proficiency requirement was reviewed by teams of elementary teachers, secondary teachers, and higher education faculty in the applicable areas (Wheeler, 1988). The dominant need for revision expressed was for some form of performance assessment, the nature of which varied with the subject-matter area. For six of the tests (Art, Biology, English Language and Literature, Industrial Arts, Introduction to the Teaching of Reading, and Social Studies), at least some reviewers identified inconsistencies with the relevant Curriculum Frameworks. In mathematics, classroom teachers and higher education faculty reviewing the relevant Specialty Area Test tended to disagree as groups about the desirable test content. Practicing teachers wanted the test to give more emphasis to topics most often taught at the intermediate and secondary level, while the higher education faculty desired more emphasis on higher level mathematics.

The NTE Specialty Area Tests are under revision and are scheduled to be replaced by new examinations in the Praxis series of assessments currently under development. The CTC has collaborated with the Praxis test developer, Educational Testing Service, in the redevelopment effort. The appropriate Praxis tests are scheduled to replace the NTE Specialty Area Tests in 1992.

The movement for "authentic," i.e., performance-based, assessment for both students and teachers led to a concern that the current subject-matter tests did not fully measure the ability to apply subject-matter knowledge in some integrated demonstration of competence such as writing an essay, solving mathematical problems, or designing an experiment. To

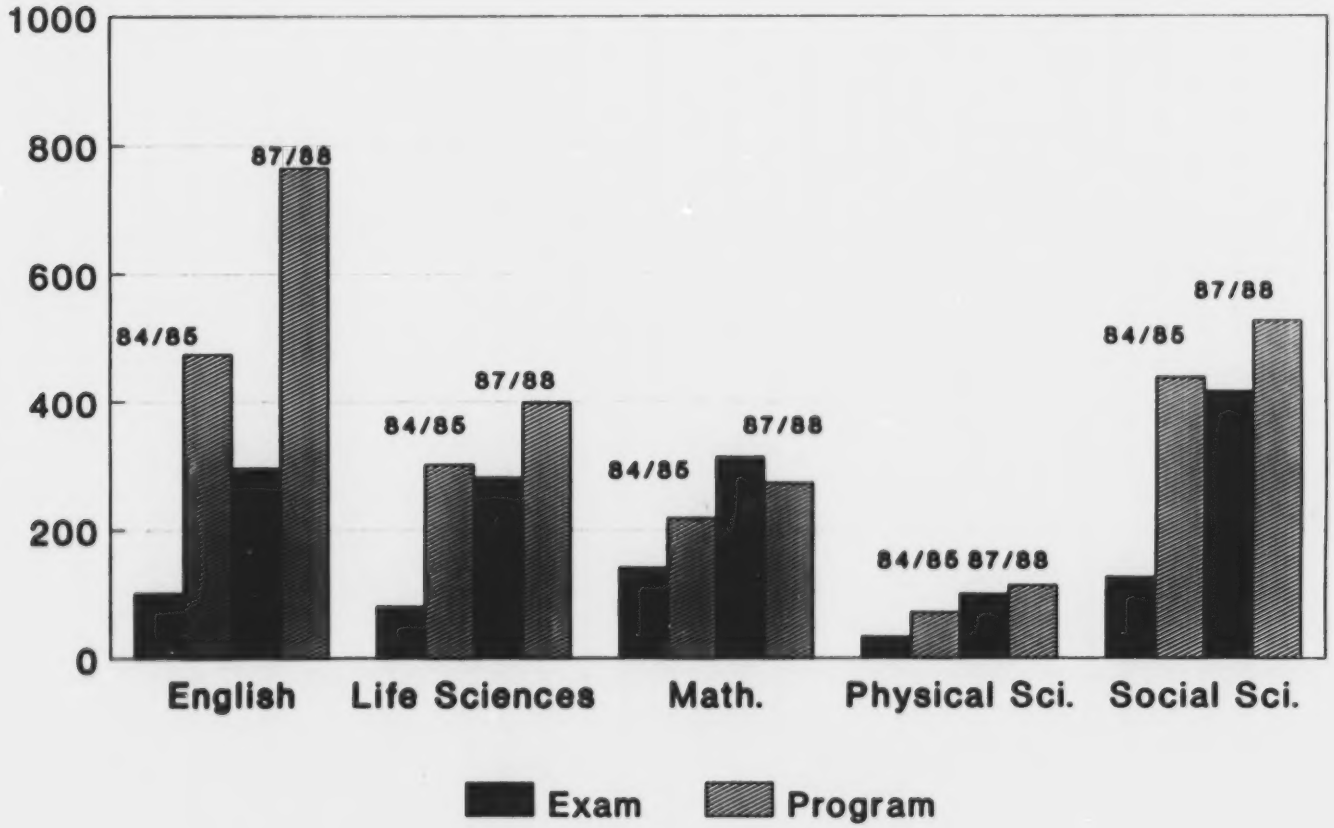
address the desire for performance assessment, the CTC authorized the development of Content Area Performance Assessments (CAPAs), to be administered in conjunction with the appropriate NTE Specialty Area Test or Praxis assessment. CAPAs for five fields (English, life science, mathematics, physical science, and social science) have been developed, and were implemented in the fall of 1991. CAPAs in the fields of art, music, physical education, French, and Spanish are in the process of development, and are scheduled to be implemented in the fall of 1992.

Trends in the use of examinations to demonstrate subject-matter knowledge. Although the legislation focussed on demonstration of subject-matter knowledge through an examination, provision was also made to waive this requirement for applicants demonstrating subject-matter knowledge through an approved program of study. Among credential applicants who prepare for teaching in California colleges and universities, around three-quarters meet the requirement for subject-area competence through this method. However, both the number and percentage of applicants who opt for examination over completion of a program has grown in the past decade. Figure 4.1 shows the number of students recommended by California programs for teaching credentials in five subject areas in 1984/85 and in 1987/88. A comparison of the two years shows an increase in the number of university recommendations in which subject-matter competence has been demonstrated by an NTE specialty exam. These data are provided by the 69 public and private universities in California which provide both subject specialty programs and professional preparation programs. The increase in the use of the NTE examination to demonstrate subject-matter knowledge is not solely due to changes in the preferences of teacher candidates, however. An increasing number of colleges and universities are requiring passage of an NTE specialty area test, whether or not an approved subject area program has been completed by the candidate. Furthermore, many teacher candidates in fields with a low demand take the NTE examination(s) in an attempt to add authorizations to their credential so as to teach in additional fields.

Review of Credential Applications

The CTC not only requires teacher candidates to exhibit specific competencies (e.g., basic skills competencies, subject-matter knowledge), but also requires completion of a teacher preparation program that meets California standards. In 1986, the Commission changed from providing guidelines for credential programs to setting standards that the programs must meet. One of the standards specifies skills to be demonstrated by program graduates. Documents submitted by a teacher candidate must indicate mastery of these skills to justify issuing a credential.

Figure 4.1: University Recommendations
CREDENTIALS BY PROGRAM AND EXAM



Reviews of an applicant's coursework and preparation are not only conducted by the CTC, but also by approved credential programs at local colleges and universities. In reviewing their own students and other applicants referred to them, teacher preparation programs are guided by standards of program quality which are set by the CTC for each type of credential program. The adherence of programs to these standards is monitored through extensive documentation of program structure and operation together with site visits. This intensive monitoring occurs every six years when a credential program's approval for recommending candidates for a teaching credential is reconsidered. In addition to the monitoring of the credential program, CTC staff also conduct a final review of any application reviewed by a local college or university.

Overview of the Major Assessment Point

A teacher's performance on the State assessments is examined by the CTC at a single point: the point at which a teacher candidate applies for a California teaching credential. This is not to say that all the assessments are administered at the same time. Individual teachers vary in the point at which they take the uniform State-administered assessments. Typically, prospective teachers applying to California credential programs complete the State-administered assessment instruments either prior to admission or shortly afterward. Graduates of out-of-state teacher preparation programs generally complete the State-administered assessment instruments prior to application for a credential. Some teachers coming from out of state are unable to take the CBEST prior to beginning employment in a district because it is only administered in California and Oregon. These teachers can qualify for a one-year nonrenewable preliminary credential, and have one year to pass the CBEST to qualify for a subsequent teaching credential.

Range of Methods Used

Currently, the CTC uses the following methods of assessing prospective beginning teachers: multiple-choice items, constructed response items, performance-based items, and document review. A given standardized test generally consists of a combination of item types. The processes and criteria used to evaluate prospective teachers vary with the focus of the assessment. Since the criteria are closely related to the focus of the assessment, they will be discussed in the next section which concentrates on the focus of assessments; the remainder of this subsection discusses the processes used to evaluate teachers in the following areas: subject-matter knowledge, general and content pedagogical skills, proficiencies in basic skills, personal attributes, and specialized knowledge.

Verification of subject-matter competence is performed by one of two methods: a standardized examination or a document review process. The 1970 Ryan Act requires passage of a subject-matter examination. At the present time, the examinations used include the NTE General Knowledge Examination (for multiple subjects teachers), and the appropriate NTE Specialty Area Test, which is further augmented by CAPAs in the subject areas of English, life science, mathematics, physical science, and social science. Eventually, portions of the Praxis series of assessments will replace the NTE Specialty Area Tests, and CAPAs will be developed for each subject area. The current set of NTE examinations employ multiple-choice questions, with a few (e.g., the foreign language areas) requiring some sort of performance assessment. When the Praxis series is adopted, all single subject candidates selecting an examination as the means of satisfying the subject matter proficiency requirement will complete a multiple-choice test, constructed response items, and a performance assessment. The decision on the format of the examination for multiple subject credential candidates awaits analysis of the field test of the Subject-Matter Assessment of Elementary Teaching to be conducted in the spring of 1992. The subject-matter examination requirement is waived for teachers completing a program of coursework meeting specific requirements. Assessment of whether or not a candidate has completed an appropriate "waiver program" is performed by either CTC staff or staff at an approved credential program who examine transcripts of coursework and sometimes interview the candidate to obtain more information about specific courses. In the latter case, the decision is reviewed by CTC staff.

A teacher's skills in general pedagogy and content pedagogy are assessed through reviewing documents for an indication of the satisfactory completion of an approved California credential program meeting specific guidelines or an equivalent out-of-state program meeting California standards for professional preparation.

Basic skills proficiencies are demonstrated by a single method: passage of the CBEST. The CBEST measures reading and mathematical skills through the use of multiple-choice items, and measures writing ability by requiring the candidate to write two essays based on prescribed prompts.

In addition to assessing the skills and knowledge described above, the State assesses personal attributes and other specialized knowledge. The personal attribute evaluated is whether or not the candidate is of a good enough character to be trusted with children. This is signified by the absence of criminal convictions or, for some convictions, evidence of rehabilitation. This is evaluated through a check of fingerprints by the Department of Justice, the FBI, and Interpol. Prospective teachers are also required to demonstrate

knowledge in two additional areas: reading methods and the U.S. Constitution. This knowledge is demonstrated by either coursework on transcripts submitted or by passage of a relevant examination. In the case of the reading course, the NTE Specialty Area Test on the Introduction to the Teaching of Reading satisfies the knowledge requirement.

Purposes and Uses of Assessments

The sole use of assessments by the State is in the issuance of teaching credentials. The purpose is to ensure that all teachers in California public schools meet minimum requirements to avoid harm to California public schoolchildren. Other organizations also use the State-sponsored assessments; these uses will be discussed in a later section.

Technical Quality

Procedures to assure technical quality differ for the review of documents and for standardized tests, ar[illegible]therefore they will be discussed separately.

Document review. Credential applications and other supporting documents are examined by certification officers or certification technicians. Certification technicians handle the less complex applications, e.g., renewals and California-trained applicants; certification officers handling the more complex applications, e.g., out-of state applications and supplemental authorizations. Certification technicians and officers examine all credential applications, not all of which are from beginning teachers. Certification officers must be familiar with 32 different types of credentials issued under the Ryan Act, the major legislation currently in effect, and 90 other credentials issued under prior or other legislation. All certification technicians and officers receive continuing updates on changes in the laws or regulations; new hires must complete a training program before they are allowed to independently process applications.

Both certification technicians and officers go through a training process, which combines aspects of direct instruction with aspects of apprenticeship. One CTC staff member, a certification officer herself, conducts the training, which lasts approximately six months for a certification technician and approximately one year for a certification officer. The length of training varies according to the length of time it takes an individual to demonstrate a high degree of accuracy in examining applications.

The certification training can occur either individually or in a group, depending on how many new people are hired. It begins with a less complex type of credential, e.g., the

thirty-day substitute, and adds on additional credential types until all have been covered with demonstrated competence in each. Trainees receive written materials, including samples of applications, which are accompanied by a review of the history and context of the development and administration of the credential, a lecture/demonstration of how to process the application, identification and explanation of reference materials, and practice in analyzing that type of application. All work of trainees is monitored by the trainer to check the accuracy of the evaluation and that all relevant information has been correctly recorded. This training is repeated for each type of credential that the trainee is to be assigned to process. Certification officers are trained to process the same types of applications as certification technicians plus an additional set of more complex types.

After demonstration of proficiency in processing an application, the work of certification technicians and officers is no longer directly checked. However, a Quality Control Unit which examines credentials before they are mailed, makes sure that each credential corresponds to the application and is issued for the appropriate length of time. In addition to the Quality Control Unit, errors are also sometimes reported by California-approved credential programs, school districts and applicants. The supervising certification officer examines both individual and collective trends in errors reported from all these sources, and determines whether additional training or changes in procedures are necessary to ensure a high level of accuracy.

Standardized tests. The standardized tests are highly monitored with respect to state-of-the-art indicators of technical quality and meet the American Psychological Association guidelines for reliability and validity. The emphasis is on decision reliability because of the nature of the test (i.e., for certification). For this reason, the focus is on reliability at the cut point (the point distinguishing passing and failing scores).

Each test for each administration is developed according to specifications adopted by the CTC. Prior to adoption, the specifications are reviewed by practicing teachers and others such as school administrators and teacher educators. Individual multiple-choice items and essay topics are also reviewed by panels for potential bias toward specific ethnic or gender groups. Potential multiple-choice items are pilot tested in each test administration. Each item's performance is then analyzed in terms of its difficulty relative to other items and with respect to differences in difficulty for examinees of different ethnic backgrounds and genders. Individual items whose pilot test performance differs in difficulty for different genders or ethnic groups are subjected to further review by a panel. Through the process of pilot testing and statistical analysis of items, comparability of scores is maintained from one form of the test to the next. This process of calibration includes

analysis of the statistical properties of entire tests as well as individual items. Essay items and performance items are piloted in separate procedures.

Multiple-choice items are machine-scored. Items which are more variable in nature and require professional judgment, e.g., writing samples or performance items, are scored by a different methodology. Each scoring session has a Chief Reader and a number of Table Leaders, who supervise small groups of scorers. The Chief Reader and Table Leaders read enough responses to collect a set for training. For each rating point, responses are identified which fit the guidelines in the scoring protocol defining that point. The guidelines maintain comparability across test administrations. The scorers, who have expertise in both teaching and the subject area, are trained simultaneously as a group to the standards, defined both by guidelines and by specific examples. Scorers practice scoring previously rated responses. After a half to two-thirds of a day of training, a proficiency test is administered. Only scorers who meet the standard, defined as a certain number of correctly scored papers, are allowed to score the actual responses. Previously scored responses are used intermittently during the actual scoring to check for and maintain reliability between scorers, i.e., that the scoring standard is maintained throughout the scoring session. Each response is scored independently by two people. If the two scores differ by more than one rating point, the discrepancy is resolved by the Chief Reader. The double scoring also serves as a method of insuring reliability of scoring. In addition, essays of candidates failing the writing portion of the CBEST are reappraised to guarantee accuracy if they either passed both the Reading and Mathematics sections or the essay score was within one point of passing. This practice helps guard against failing a person who might otherwise have passed due to scoring errors in the judgment-based scoring portion of the examination.

Focus of Assessments

The general foci of the assessments have been described earlier. This section presents more specific detail about the following foci: subject-matter knowledge, general and content pedagogy, and basic skills proficiencies.

Subject-matter Knowledge

Under the current Ryan Act, a teacher is required to demonstrate subject-matter knowledge by passing an NTE test and, for some subject areas, a CAPA examination. As an alternative, the teacher can complete an approved course of study (i.e., set of courses) addressing specific topics. Most prospective teachers prepared in California satisfy the

subject-matter requirement through the alternative coursework, whose structure emphasizes breadth of preparation across the subject area with some more specialized coursework to reflect depth of knowledge in some area.

Altogether, 16 NTE Specialty Area Tests plus the General Knowledge section of the NTE Core Battery are currently authorized for use in credentialing by the State. Fifteen of these correspond to the 15 single subjects for which California teaching credentials can be obtained. These are:

- Art Education
- Agriculture
- Biology and General Science
- Business Education
- Technology Education
- Chemistry, Physics, and General Science
- English Language and Literature
- French
- German
- Home Economics Education
- Mathematics
- Music Education
- Physical Education
- Social Studies
- Spanish

The NTE Specialty Area Test on the Introduction to the Teaching of Reading can substitute for a course on the teaching of reading, which is required of all credential applicants in California. The NTE Specialty Area Tests not only cover subject-matter knowledge but also knowledge of general pedagogical principles. The test specifications for each subject are developed in consultation with practicing teachers to reflect the emphasis placed on skills in the field. Many of the tests not only address knowledge of concepts and principles but also knowledge of current developments in curriculum, important subject-specific professional organizations, and the history of the discipline.

In five subject areas (English, life science, mathematics, physical science, and social science), the NTE Specialty Area Test is accompanied by a Content Area Performance Assessment (CAPA.) The CAPAs focus on the ability to apply subject-matter knowledge. While application is still addressed through multiple-choice items in the NTE Specialty Area Tests, the CAPAs use a different format to require some sort of performance to demonstrate the ability to apply knowledge. CAPAs for art, music, physical education, French, and Spanish are in the process of development and are expected to be in place in the fall of 1992.

Multiple subject credential applicants satisfying the subject-matter knowledge requirement through examination currently do so by passing the NTE General Knowledge Examination. However, that examination will soon be replaced by a test under development, which is tentatively titled the Subject Matter Assessment of Elementary

Teaching. This test design closely parallels the subject-matter knowledge requirements for candidates satisfying the requirement through designated coursework. The current design being piloted is as follows: The test is administered in three sections, each of which requires two hours to complete. The first section consists of 150 multiple-choice questions evaluating breadth of knowledge and skill in seven disciplines, at a level of difficulty comparable to lower-division undergraduate courses. The seven disciplines addressed are English/Language Arts, Human Development, Mathematics, Physical Education, Science, Social Studies, and Visual and Performing Arts. The second section consists of 22 short constructed-response items, distributed across the seven disciplines. These items assess breadth of thinking skills and knowledge which is not well evaluated by multiple-choice questions. The third section consists of two sets of constructed-response items addressing more advanced thinking skills at the upper division level of difficulty. The first set consists of one to three items which are relevant to all seven disciplines, and which are designed to evaluate the ability to integrate across disciplines. The second set consists of one or two items which are specific to the discipline that the individual candidate has selected as a subject of concentration.

The NTE Specialty Area Tests are scheduled to be phased out in the fall of 1992 and replaced by a series of examinations within the Praxis Series: Professional Assessments for Beginning Teachers. These examinations focus solely on subject-matter knowledge, and take the form of multiple-choice tests plus two optional components: one with constructed response answers and one with a performance assessment. The CTC and ETS have agreed to collaborate in the redesign of the multiple-choice and constructed-response modules, with one Commission appointee on each national committee overseeing the development of a subject-specific test.

General Pedagogy and Content Pedagogy

These assessments are performed by teacher preparation programs and reviewed by the CTC staff. Program standards require all California teacher preparation programs to confirm that a prospective teacher has developed competence in eleven areas, which are described as follows (CTC, 1990):

> Student Rapport and Classroom Environment. Each candidate establishes and sustains a level of student rapport and a classroom environment that promotes learning and equity, and that fosters mutual respect among the persons in a class.

Curricular and Instructional Planning Skills. Each candidate prepares at least one unit plan and several lesson plans that include goals, objectives, strategies, activities, materials and assessment plans that are well defined and coordinated with each other.

Diverse and Appropriate Teaching. Each candidate prepares and uses instructional strategies, activities and materials that are appropriate for students with diverse needs, interests and learning styles.

Student Motivation, Involvement and Conduct. Each candidate motivates and sustains student interest, involvement and appropriate conduct equitably during a variety of class activities.

Presentation Skills. Each candidate communicates effectively by presenting ideas and instructions clearly and meaningfully to students.

Student Diagnosis, Achievement and Evaluation. Each candidate identifies students' prior attainments, achieves significant instructional objectives, and evaluates the achievements of the students in a class.

Cognitive Outcomes of Teaching. Each candidate improves the ability of students in a class to evaluate information, think analytically, and reach sound conclusions.

Affective Outcomes of Teaching. Each candidate fosters positive student attitudes toward the subjects learned, the students themselves, and their capacity to become independent learners.

Capacity to Teach Cross-culturally. Each candidate demonstrates compatibility with, and ability to teach, students who are different from the candidate. The differences between students and the candidate should include ethnic, cultural, gender, linguistic and socioeconomic differences.

Readiness for Diverse Responsibilities. Each candidate teaches students of diverse ages and abilities, and assumes the responsibilities of full-time teachers.

Professional Obligations. Each candidate adheres to high standards of professional conduct, cooperates effectively with other adults in the school community, and develops professionally through self-assessment and collegial interaction with other members of the profession.

The records of each applicant are examined to determine whether or not these standards have been met. The quality of the evaluations by California-approved credential programs is assured by the credential program review and approval process. These reviews include interviews with graduates, cooperating teachers, and employers of graduates asking which graduates of the program have actually acquired the skills listed above.

The applications of applicants who received their teacher preparation outside California undergo a more lengthy review to determine if the teacher preparation program meets California standards. If that is the case, then the program evaluations of teaching competence are reviewed to determine whether or not the teacher has successfully mastered the requisite skills in general and content pedagogy.

Basic Skills Proficiencies

Basic skills proficiencies are assessed by the CBEST, which is a three-part examination, with one part corresponding to each of the three areas of basic skills mandated. The reading section consists of 50 multiple-choice questions to be completed in 65 minutes. The math section comprises 50 multiple-choice questions to be completed in 70 minutes. The writing section contains two essay topics, to be completed in 60 minutes. In the two multiple-choice sections, 40 items comprise the test proper, while the remaining ten items are included for pilot testing purposes. Raw scores on each of the three sections are converted to a 20 - 80 point scale for reporting.

The **Reading Section** of the CBEST is intended to test three forms of reading comprehension: 1) literal comprehension (25% of the total items); 2) inferential comprehension (50%); and 3) critical comprehension (25%). The weights given to the three comprehension skills are approximate; any given item may draw on two or more of them. "Literal comprehension" means to be able to recognize the main idea of a passage or to identify explicit details, facts, ideas or other information presented in the passage or that is presented in a chart or graph. "Inferential comprehension" refers to the ability to identify inferences or conclusions that, though not explicitly stated in the passage, can be reasonably drawn from and supported by information explicitly presented there. "Critical comprehension" requires the reader to make judgments with respect to the style,

organization, or logic of the passage. Appendix A provides sample items illustrating these three forms of comprehension. In the complete test battery, reading comprehension questions refer to passages of one sentence, one paragraph, or three or four paragraphs in length, as well as to charts and tables.

The **Mathematics Section** of the CBEST, like the Reading Section, consists of 50 multiple-choice items. The items are designed to test three constructs of mathematical ability: 1) problem solving processes; 2) solving applied problems; and 3) knowledge of mathematical concepts and relationships. Various items test these abilities as they apply to arithmetic (up to 40% of the total items), geometry (at least 20%), and algebra (the remainder). "Problem solving processes" test knowledge of how to go about solving a problem, e.g., identifying the operation (such as addition, multiplication) needed to solve a problem. "Solving applied problems" means the ability to solve fairly straightforward word problems. "Knowledge of mathematical concepts and relationships" tests the understanding of basic concepts, such as the meaning of certain terms, order among numbers, relationships shown by graphs, elementary probability, and the like. Appendix A shows sample items testing these three constructs as they apply to arithmetic, geometric, and algebraic content.

The **Writing Section** of the CBEST, in contrast to the Reading and Mathematics Sections, elicits a performance, in the form of a writing sample, from the examinees. The Writing Section tests the candidate's ability to write effectively in a limited period of time. It consists of two topics, conveyed in one or two sentences, on which the examinee must write short essays. One topic asks for an analysis of a situation or statement; the other asks the examinee to write about a personal experience. The writing is scored holistically on the basis of such things as the quality of insight or central idea, evidence of knowledge of the audience for whom the piece is being written, clarity, consistency of point of view, cohesiveness, strength and logic of supporting information, rhetorical force, appropriateness of diction and syntax, and correctness of mechanics and usage.

Differential Impact of Assessment Practices

While IHEs and LEAs typically do not systematically collect information on the impact of their assessment practices on various groups, comparisons are performed and analyzed for every standardized test administered by the State. The CTC staff's annual report to the Commission on the results of the CBEST examination includes profiles of examinees and passing rates of various groups of interest. As discussed earlier in the section on technical quality, empirical bias reviews are conducted after each test administration for both the CBEST and the NTE Specialty Area Tests.

Ethnic groups. Differences in passing rates between ethnic groups have persisted since the inception of the CBEST (See Figure 4.2.) Passing rates on the CBEST have increased over time for all ethnic groups. In 1982/83, the year in which the CBEST was first administered and passing standards established, 68% of all examinees passed the test. In 1989/90, 74% of all first-time examinees passed the entire test. Passing rates of first-time examinees within all ethnic groups have increased between 1982/83 and 1989/90, the latest year for which data is available. Over this time period, the passing rate for Hispanics taking the exam for the first time increased from 39% in 1982 to 50% in 1989/90, for Blacks from 26% to 38%, for Asians from 50% to 62%, and for Whites from 76% to 81%.

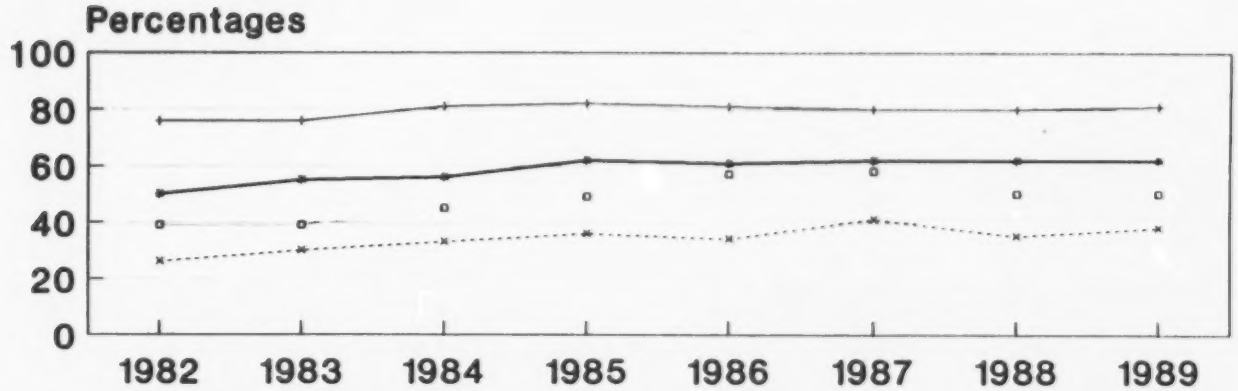
Since examinees are allowed to retake the CBEST if they fail one or more sections, the ultimate concern about ethnic differences in passing rates is about differences in *cumulative* pass rates. A 1990 study (Majetic, 1990) examined cumulative pass rates for different ethnic groups. The study focussed on the 1985 cohort of test-takers. Differences between ethnic groups in cumulative passing rates, reflected in Table 4.1, persist, but are smaller than differences in initial rates. These cumulative passing rates were affected by the difference among ethnic groups in the percentage of those who did not pass who elect to retake the examination. While rates for Hispanics and Whites are similar (roughly 75%), substantially lower percentages of Asians and Blacks elected to retake the CBEST after failing (60 and 52% respectively.)

TABLE 4.1

INITIAL AND CUMULATIVE CBEST PASSING RATES FOR EXAMINEES WHO ORIGINALLY TOOK THE TEST IN 1985 BY ETHNIC GROUP MEMBERSHIP

	Asian Americans	Black Americans	Hispanic Americans	Caucasian Americans
Initial Passing Rates in 1985	62.3	36.1	50.7	82.2
Cumulative Passing Rates: 1985-90	83.7	63.2	86.3	96.4

FIGURE 4.2
CBEST PASS RATES, 82/83-89/90



	1982	1983	1984	1985	1986	1987	1988	1989
Total	68	69	73	75	75	75	74	74
White	76	76	81	82	81	80	80	81
Asian	50	55	56	62	61	62	62	62
Hispanic	39	39	45	49	57	58	50	50
Black	26	30	33	36	34	41	35	38

White — Asian ° Hispanic ··· Black

Source: Watkins, 1985b, p. 14 and Majetic, 1990, p. 15

4.19

An examination of passing rates of NTE Specialty Area tests by ethnic group membership shows patterns similar to those seen in the CBEST. For most exams, passing rates tend to be highest among Whites, followed by Asians, Hispanics, and Blacks. For the English subject test, for example, in 1989/90 53% of White examinees passed, compared to 29% of Asians and Pacific Islanders, 28% of Hispanics, and 7% of Blacks. The corresponding figures in the Social Studies test are 60% for Whites, 40% for Asians and for Hispanics, and 27% for Blacks. In the Mathematics and Life Sciences tests, however, pass rates for Asians exceed those of Whites, with 54% of Asians passing in Life Science and 35% in Mathematics compared to 48% of Whites passing in Life Sciences and 34% in Mathematics.

Gender groups. There are also marked gender differentials for some NTE Specialty Area Tests. In Art, Physical Education and the Introduction to the Teaching of Reading, the pass rates for females exceed those of males by 5%, 9% and 10%, respectively. In English, Spanish, and Mathematics, pass rates for females are 2% to 3% lower. The largest differences are seen in Life Science, with 37% of females and 52% of males passing; Physical Science, 32% of females and 51% of males; Music, with 44% of females and 76% of males, and Social Studies, with 52% of females and 61% of males passing.

Type of credential. Passing rates for the NTE also vary between subject areas. In 1989/90, these rates ranged from a low of 33% of the 1,956 examinees in mathematics to a high of 92% in agriculture, in which only 27 individuals took the exam. Among the subjects with larger numbers of examinees, the passing rates were 51% of 1,540 in English; 44% of 931 in Life Sciences; 46% of 462 in Physical Science; and 58% of 1,925 in social studies.

The CBEST has a differential impact across types of credentials, illustrated in Table 4.2. The group of first-time examinees seeking single subject credentials with no additional emphasis has a slightly higher pass rate than those seeking multiple subjects credentials with no additional emphasis. However, prospective bilingual teachers, at both elementary and secondary levels, have greater difficulty with the CBEST compared to other groups taking the test for the first time.

Review of studies of potential sources of differential performance. The use of the CBEST as a credentialing requirement has further reduced the already small pool of potential teachers from nonwhite ethnic groups. In order to understand more fully the reasons for these differences, the CTC has commissioned several studies to investigate possible causes, including the effect of time limits and potential bias in item content.

TABLE 4.2

CBEST PASSING RATES FOR FIRST-TIME EXAMINEES
GROUPED BY CREDENTIAL BEING SOUGHT

Credential to be Applied For	1986–87		1987–88		1988–89		1989–90	
	Number Tested	Percent Passing	Number Tested	Percent Passing	Number Tested	Percent Passing	Number Tested	Percent Passing
Multiple Subjects	6,198	71%	5,016	73%	5,703	69%	5,673	71%
Multiple Subjects with Bilingual Emergency	595	52%	307	54%	587	42%	684	45%
Single Subjects	4,952	77%	2,900	77%	4,020	76%	3,820	76%
Single Subjects with Bilingual Emergency	301	51%	150	52%	305	54%	298	52%
Emergency Teaching	4,196	78%	2,236	82%	3,213	78%	3,721	80%

Source: Majetic, 1990, p. 12

The differential performance of groups has been a major concern, so several special studies have been sponsored by the Commission to explore possible sources of the differences in performance unrelated to the skills and competencies being measured. A 1984 study of the effects of time limits in the CBEST studied statistical comparisons of clusters of items placed at different points in the test. Minor modifications in the test were made as a result of the study, but the time limits per se did not appear to explain differences in group performance.

A second potential source of bias that has been subjected to study relates to the subject matter of test items. Analysis for such bias is a standard step in the process of constructing test items, each of which is reviewed by a five-member group. In addition, expert groups have examined several forms of the CBEST for potential ethnic or gender bias, in terms of stereotypic or potentially offensive items. Members of these panels identified few individual items which appeared to carry bias against particular ethnic groups or women. A statistical analysis of these items showed no correlation between performance on these items and membership in an ethnic group that might perceive the item as biased.

After each test has been given, an empirical bias review is conducted. Each item is examined in terms of comparative difficulty for various subgroups. The group comparisons made are male/female, white/black, white/hispanic, and white male/white female. Items which are above a threshold index of difficulty are then individually examined for bias by a panel of test specialists, with the panel required to give a rationale for the continued use of items with extremely large group differences. Individual items which act differentially by gender or ethnic group are subjected to review by a panel. Although item-level analysis has provided little insight into differences in performance between members of different ethnic groups, it pointed to some difficulties specific to non-native speakers of English. For these examinees, interpreting passages that rest on double meanings, or understanding questions formulated in the negative (e.g., "Which of the following does the author not claim?") seems to pose particular problems.

The cumulative pass rate study of the CBEST (Majetic, 1990:9) does identify strengths and weaknesses which contribute to differences in passing rates. As a group, Blacks exhibit markedly higher levels of skill in reading and writing than in mathematics. Asian examinees are weak in reading and writing in English relative to their mathematics skills, while Hispanic candidates as a group are weak in writing and mathematics compared to reading.

Costs

In 1990-91, it cost prospective teachers $35 to take the CBEST; repeat examinees who had passed the writing portion were eligible for a refund of $3.25. It cost prospective teachers $60 for each test to take the NTE Specialty Area Tests and the CAPAs.

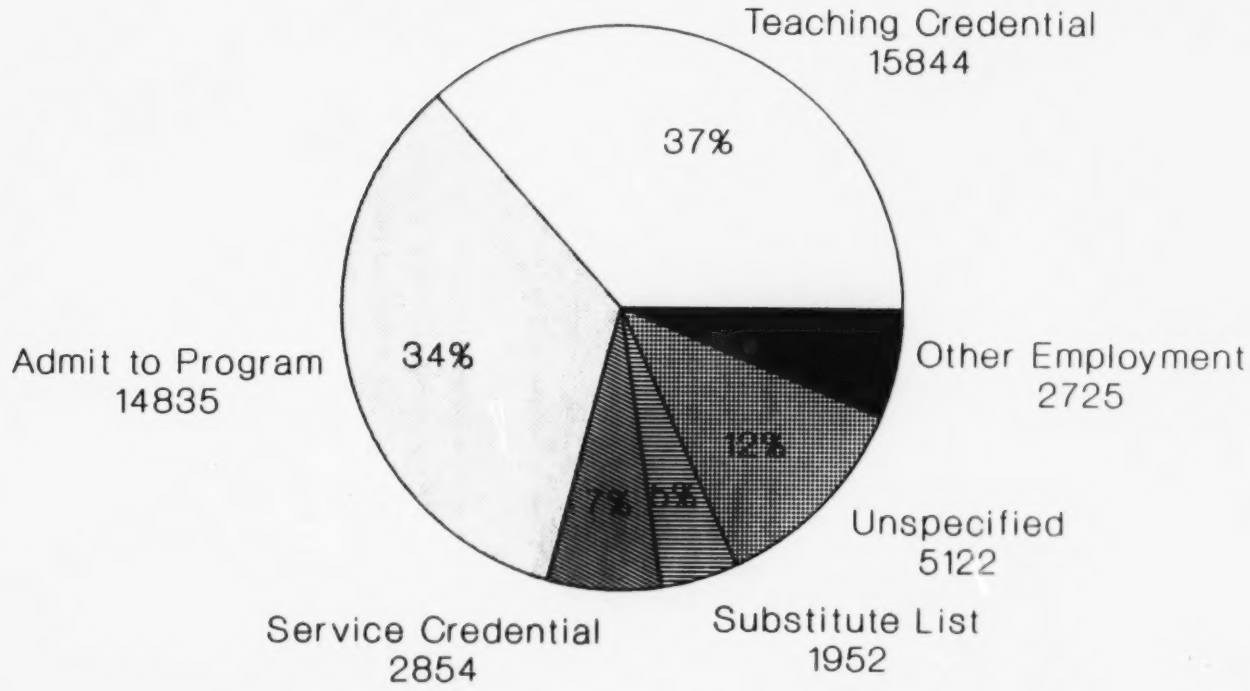
Contribution of State Assessments to IHE and LEA Decisions

IHE Decisions

The IHEs use the CBEST at various points. Applicants to California credential programs are required by law to take the CBEST prior to admission to the program for diagnostic purposes and to pass it prior to student teaching. Some California IHEs use passage of the CBEST as a screening requirement for admission. Other campuses, including campuses whose service area includes large numbers of persons from groups who tend to fail the CBEST in higher proportions than average, require their students to *take* the CBEST prior to admission, but they don't have to pass it at that point. Some of these campuses provide courses which are directed at improving student performance on CBEST. Students who are admitted despite failing the CBEST must pass it prior to the commencement of student teaching, as required by law. Figure 4.3 illustrates the distribution of reasons for taking the CBEST among first-time 1989-90 examinees, indicating that at least 34% were taking the CBEST for admission to a credential program.

Although the reason for requiring the CBEST prior to admission to a credential program is for diagnostic purposes, the diagnostic information provided by CBEST results is limited. The CBEST has too few items per skill to serve a detailed diagnostic function, although the probability of passing the CBEST on a subsequent attempt with no intervening instruction is projected. A Basic Academic Skills Test, which is scheduled for release in 1992 by ETS, includes an optional computer-delivered diagnostic/practice component similar to, but longer than, the written test. The instructional modules accompanying the diagnostic/practice test will deliver instruction for the instructional units diagnosed as weak. This test was not developed solely with California specifications in mind. However, if it is found to satisfy the same function as the CBEST, the diagnostic component could assist individual examinees as well as the institutions of higher education who recruit from underrepresented populations and provide instructional assistance to those students who fail the CBEST.

FIGURE 4.3
First-time Examinees' Reasons for Taking CBEST, 1989/90



Source: Majetic, 1990, p. 11

Passage of a relevant NTE and possibly a CAPA is required prior to student teaching unless applicants have completed a program of coursework which meets the criteria for waiving the requirement. Less than half of the credential applicants use an examination as the means for satisfying the subject-matter requirement.

LEA Decisions

Passage of the CBEST is required for a teaching credential which enables a teacher to be eligible for a regular teaching position. The only exception is for out-of-state applicants who can delay passing the CBEST for one year if the employing district is having difficulty finding teachers and requests a one-year nonrenewable teaching credential for the applicant. The NTE is also required for a credential unless the applicant has completed a waiver program. Generally, NTE scores were not mentioned as a factor in LEA decisions by the district informants in the LEA substudy.

Summary and Conclusions

The State assessment practices substudy examined the evaluation of beginning teachers at the State level. This section summarizes the findings and conclusions.

Assessment Context

State assessment practices conform to legislative requirements and test specifications developed in consultation with teachers, administrators, and teacher educators. Curricular changes have recently led to the revision of subject-matter examinations, which either do or will shortly include performance-based measures.

Methods of Assessments

Methods of assessment of beginning teachers include document review and standardized tests. The standardized tests use one or more of the following types of items: multiple-choice, constructed response, and performance-based.

Focus of Assessments

The State assessments focus on the following areas: subject-matter competence, general pedagogy and content pedagogy, basic skills proficiencies, personal character, and knowledge of reading methods, and the U.S. Constitution. The assessments are used in

issuing teaching credentials; the purpose is to ensure that all teachers in California public school meet minimum requirements.

Subject-matter knowledge tests concentrate on knowledge of concepts and principles and current developments in curriculum, and in some areas are augmented by performance assessments which assess the ability to apply subject-matter knowledge. The specific topics and skills are determined for each subject area by teachers and teacher educators, so they vary from area to area.

The evaluation of a prospective teacher's knowledge of general pedagogy and content pedagogy addresses skills prescribed by the State, but is performed by the credential program from which they graduated. The State confirms that the requisite skills were evaluated favorably by the program. Proficiency in reading, writing, and mathematics at a basic collegiate level is required to be demonstrated through passage of a designated test, the CBEST.

Technical Quality of Assessments

The technical quality of standardized assessments administered by the State is very high. These assessments meet all APA standards for reliability and validity. Items are pretested and analyzed to maintain comparability across different administrations of the same test. Even procedures for document reviews are highly standardized. Both in document review and in scoring of tests where evaluation requires professional judgment, evaluators are trained following a detailed protocol and monitored to maintain reliability across evaluators.

Although differences in passing rates have been documented and continue to exist for some groups, several studies have failed to identify any characteristics of the assessment instruments which produce such differences that are unrelated to differences in the skills being measured.

Assessment and Support

At this time, standardized State assessments are designed to be reliable at the decision point and are not designed to yield diagnostic information. These tests yield a pass/fail decision reflecting performance across a number of dimensions and topics, and will not provide diagnostic information for individual examinees, institutions of higher education, or districts wishing to know more details. More diagnostic information about

basic skills proficiencies would be especially useful for IHEs seeking to address the underrepresentation of specific groups in teaching. Both IHEs and LEAs might find more information about a teacher's subject-matter knowledge useful.

Conclusions

Generally, the State assessment practices meet higher standards of technical quality than the IHE and LEA assessments. Great care is taken to maintain uniformity in the administration and evaluation of assessments. Since the emphasis is on the decision as to whether or not to issue a credential to each applicant, resources have been devoted to precision at the decision point. Although providing diagnostic information is not necessarily a mutually exclusive goal, at this time, only a limited amount of information is available to guide any additional preparation for candidates who fail the standardized State assessments.

CHAPTER 5:

EXPERIENCES OF BEGINNING TEACHERS

CHAPTER 5:

EXPERIENCES OF BEGINNING TEACHERS

This report presents the findings of a study of 22 beginning teachers currently employed in the state of California. For the study, an interview of one to three hours in length was conducted with each of the teachers about their experiences being assessed at seven different points during their education and employment. In addition, each teacher consented to the release of assessment information from their teacher preparation program and school district files.

The report is divided into four sections. First is a description of the study sample, including the teacher participants and the IHEs and LEAs that they represent. Following this is a discussion of the assessments that the teachers actually experienced as they "traversed" the road to becoming a credentialed and beginning teacher. The third section discusses the content of the assessments, based on the teachers' perceptions. Concluding the report is a summary of findings culled from the teacher interviews.

Sample Description

Of a sample of 25 identified beginning teachers, 22 eventually participated in the study (3 teachers withdrew at various points). Some characteristics of the teachers are presented below, followed by a description of the range of institutions of higher education (IHEs) and California school districts -- referred to as local education agencies (LEAs) for this report -- that the teachers represent.

Teachers

Although the 22 teachers selected for this study are different in many ways, they share the following characteristics:

- Employed as beginning teachers by California school districts since 1988-89;
- Participated in FWL's pilot studies of alternative assessments during Phases I and II of the CNTP Assessment Component; and,

- Provided informed consent for FWL to contact their IHEs and LEAs to obtain information about assessments of their work.

In addition, all but one of the teachers possesses at least one clear or preliminary California teaching credential. (One teacher is on an emergency credential.) The majority of these credentials are single subject credentials in the subject areas of math, science, English, and social studies--the same areas covered by the assessments piloted in Phases I and II. The numbers of teachers who have each type of credential are as follows: English (5); social science (4); science (3); and math (1). Those teachers not credentialed in these four areas, hold a multiple subjects credential.

Other characteristics of the teachers in the sample are depicted in Table 5.1. As the table shows, a slight majority of the teachers are female and teach (or taught during their first year) at the elementary level. Northern and southern California are almost equally represented, with three teachers located in the central valley. A little over a third of the teachers are minorities, and slightly more than a fourth are older than 35 years of age. Seventeen of the 22 teachers are in their second or third year of teaching.

As mentioned above, all of the teachers had participated in at least one of the pilot studies of the alternative assessments during Phases I and II of the assessment component. Of the 12 alternative assessments piloted over the past two years, nine are represented by teachers in this study. These nine assessments and the number of teachers who took them (some teachers participated in more than one assessment) are as follows:

Secondary English Portfolio	(6)
Secondary English Assessment	(6)
Semi-Structured Interview -- Secondary Social Science	(4)
Classroom Competency Instrument	(3)
Semi-Structured Interview -- Elementary Math	(3)
Science Laboratory Assessment	(3)
Structured Simulation Tasks for Secondary English Teachers	(2)
Secondary Life/General Science Teacher Assessment	(2)
Semi-Structured Interview -- Secondary Math	(1)

TABLE 5.1

BEGINNING TEACHER CHARACTERISTICS

(Total Number of Beginning Teachers Interviewed = 22)

Descriptive Characteristics of Participants	Distribution of Participants
Gender	
Male	9
Female	13
Teaching Level	
Elementary School	9
Middle School	6
High School	7
Location of Teacher	
Northern California	9
Southern California	10
Central Valley	3
Ethnicity	
Asian–American	2
Black	2
Hispanic	4
White	14
Age	
Under 35	17
35 and over	5
Teaching Experience	
First Year	2
Second Year	8
Third Year	9
Fourth Year	3

A comparison of the results of the alternative assessments with the results of the assessments made of teachers during student teaching and the first year of teaching can be found in Appendix B.

Range of IHEs and LEAs Represented

Exactly half of the 22 teachers received their teacher preparation training at a California institution of higher education, and the other half received their training either at one of the state's private schools or at an out-of-state institution. (One teacher attended an intern program.) Of the 11 teachers who attended a California IHE, 10 attended a CSU program, and one attended a UC program. (Interestingly, half of the 10 CSU teachers attended a UC campus for their undergraduate work.) The 11 public IHEs represented by the teachers are evenly distributed between northern and southern California.

Of the 11 teachers in the sample who did not receive training through a CSU or UC program, seven attended a private teacher education program in the state (six of whom also attended a private school for their undergraduate work), and three were trained out-of-state. The private schools represented in the study include institutions with religious affiliations, a very elite, high stature school, and a school with satellite campuses across the state. The three out-of-state programs are located in Arizona, Colorado and Iowa.

Like the IHEs, the LEAs represented in the study are evenly distributed among northern and southern California. They are also located in a variety of rural, suburban, urban and inner city settings. The enrollment size of the LEAs varies greatly, ranging from an elementary district with 65 students to a K-12 district with an enrollment of 592,273 students.

Traversing the Credentialing Assessment System

For the purpose of this report, we will describe California's credentialing assessment system as consisting of seven distinct points at which a prospective or beginning teacher's knowledge, skills, and/or abilities are assessed: (1) subject matter preparation in college, (2) admission to a teacher education program in a school of education, (3) education courses prior to and during student teaching, (4) admission to student teaching, (5) student teaching, (6) employment (i.e., hiring), and (7) first-year retention. The assessments at each point, as they were experienced by the teachers, are discussed below.

(Before beginning the discussions, we would like to make two notes both of which were mentioned in the methodology section but which are worth repeating here. First, all of the teachers who were interviewed are still teaching. Teachers who may have left the profession during student teaching or after were not interviewed. Thus, there is a potential bias in the sample responses. Second, although the following discussions are based on the teachers' recollections of their assessment experiences, an examination of the teachers' university and district evaluations that we have on file supports their recollections.)

Subject Matter Preparation in College

The first assessment point usually experienced by a prospective teacher is that of showing competency in a certain subject matter area or areas. This competency is needed in order to qualify for a teaching credential, whether it be single subject or multiple subjects. Subject matter competency is generally ascertained in one of two ways: (1) passing an NTE exam, or (2) completing subject matter courses in college.

The majority of the teachers in our sample took and passed an NTE exam, while eight teachers qualified for a credential by completing subject matter courses in college. Each of these eight teachers was asked questions about the assessments they experienced in their subject matter courses. Since almost all of the teachers had completed the courses several years ago, it was, perhaps, not surprising that they could only vaguely remember their courses and especially the kinds of assessments they experienced in the courses. The majority of teachers, however, recalled that most or all of their subject matter courses had final and midterm written exams, as well as some required writing assignment(s). Approximately half of the teachers could not recall their courses having any other types of assessments, such as required demonstrations or performances, and the other half said their experience with such assessments was minimal.

The teachers' recollections of the format of their final exams in their subject matter courses match the data provided by the IHEs--i.e., most exams were a combination of essay and multiple-choice questions, or essay questions alone. In terms of difficulty, all of the teachers said they had to study hard to earn a B or better on <u>most</u> or <u>all</u> of these exams.

Aside from the final examinations in the subject matter courses, almost none of the teachers said they experienced any further assessment of their subject matter knowledge at this point. Only one teacher said she had to pass a comprehensive exam to earn a Bachelor's degree, and she described this exam as a senior thesis. This teacher completed her coursework at an Ivy League college.

Admission to a Teacher Education Program

The second assessment point which served as a focus of discussion in the teacher interviews was that of admission to a teacher education program in a school of education. The discussion began by asking the teachers the number of teacher education programs to which they applied. Approximately 80% (17 of 21) of the teachers applied to only one program. At the most, one teacher applied to three programs.

The teachers were then asked to recall the criteria or requirements they had to meet to get admitted to their program. (If not mentioned spontaneously, interviewers probed for possible criteria.) Most teachers recalled that they had to have a minimum GPA, complete a subject matter program, pass a test (the CBEST was cited most often), and write an essay (usually on the topic, "Why I Want to Teach"). Less than half the teachers recalled having to be interviewed by an admission person/committee, having to complete specific prerequisite courses, or needing experience with children. In fact, only 4 of the 21 teachers recalled needing experience with children. Although some of the teachers who experienced admission interviews said they were asked then about their experience with children (most IHE applications for admission into the teacher education programs have at least one question that asks about experience with children), less than 20% of the teachers seemed to perceive that experience as a requirement for admission into their teacher preparation program.

All but two of the 21 teachers said they met all of the requirements to get into their teacher education program. One teacher had a requirement waived, and another was put on probationary status because of her GPA. With such a high success rate, it is perhaps not surprising that only a third of the teachers definitely considered the screening process for their program to be rigorous. Some of these teachers cited their program's GPA requirement, while one teacher told of a "rigorous interview process." Almost 50% (10 of 21) of the teachers were of the opinion that their program's screening process was definitely not rigorous, and the remaining teachers were undecided. One of the teachers who did not find his program's screening process to be rigorous based much of his opinion on the types of people he encountered in the program after it started, recalling "English teachers who couldn't write and biology teachers who made elementary errors."

Education Courses Prior to and During Student Teaching

As part of their teacher education programs, all of the teachers took some education courses, the majority of which were taken either prior to or during student teaching. The teachers were asked various questions about the assessments in these courses--questions that were very similar to those asked about the assessments experienced by the teachers in their undergraduate subject matter courses.

Although many teachers could not remember the exact number of education courses they had taken, the approximate estimate was nine courses, with a range between 4 and 15. As was the case with the subject matter courses, almost all of the teachers said that at least half of their education courses had final and midterm written examinations, as well as required writing assignments.

In addition to the more traditional forms of assessment, approximately 20% of the teachers said that all of their education courses included some demonstration or performance assessment, and all of the teachers recalled at least one course that had such an assessment. The demonstration or performance assessments most commonly experienced by the teachers were those of 1) developing a lesson plan, and 2) presenting a lesson to their peers (either the whole class or in small groups). Other demonstration/performance assessments experienced by the teachers included giving a speech on an educational topic, reading a book aloud to peers in the class (a reading class assessment), writing case studies of students, and creating portfolios. Two teachers recalled still other forms of alternative assessments: one was videotaped teaching a lesson, and the other described a nine-week project in which she was expected to take the role of the teacher in such simulated classroom activities as recording and averaging grades, adding and dropping students, and sending out memos.

While the more traditional forms of assessment were those most frequently experienced by the teachers in their education courses, it was the demonstration and performance assessments that almost all of the teachers described as most useful or helpful. In fact, only two teachers described the final or midterm written exams as most useful, both explaining that such assessments allowed them to show "the knowledge acquired."

Although the final written exams in the education courses were said by the teachers to be like those in the subject matter courses--i.e., usually composed of essay questions or a combination of essay and multiple-choice items--the level of difficulty appeared to be different. While all nine teachers who qualified for a credential by completing subject

matter courses said they had to study hard to earn a B or better on all or most of their final exams, only two of these teachers said the same thing about the finals in their education courses. The other seven teachers said they had to study hard for only some or few of their education course finals. These nine teachers, however, were not representative of the entire sample of teachers, half of which stated they had to study hard for all or most of the final exams in their education courses and the other half said they had to study hard for only some or few of their exams.

Admission to Student Teaching

While some teachers said that their student teaching was concurrent with their education courses, the majority of teachers began student teaching after they completed a block of education courses. Thus, we interviewed teachers about the assessment process they recalled experiencing in order to be admitted to student teaching. In particular, we focused on their recollection of the criteria or requirements they needed to meet before being admitted to student teaching.

More than half of the teachers recalled having to complete an observation/participation assignment and specific education courses, as well as needing a minimum GPA (usually a B or better) in those courses. Less than half (30%) said they had to go through an interview, either with someone from the university program (e.g., department director, university supervisor, bilingual advisor) or with someone at the school site (e.g., principal, department chair, cooperating teacher) where they were assigned to student teach.

The majority of teachers perceived these "assessments" to be useful, fair, and relevant. The coursework and the observation assignment, in particular, were thought to be useful because, as one teacher said, "they gave us background in giving lessons and working with students." Another teacher, who had been a speech and theater major and started teaching in his thirties, commented that these requirements were useful because before meeting them "I had no idea what was involved in teaching." Similarly, a younger teacher who had majored in science explained,

> *When you first start you have no idea what you're getting into. You need some background. A lot of people dropped out after the observation requirement.*

Of the few teachers who did not perceive the requirements as being relevant, the reason most often given was that the courses and even the observation assignment were not like "a real teaching situation" or "had nothing to do with the school/class I'm [teaching] in now." One teacher also cited the irrelevancy of his program's GPA requirement, stating that, "lots of people with low GPAs were better teachers than those with high GPAs."

While most teachers described the assessment process for admission to student teaching as useful, fair and relevant, almost none of the teachers described it as difficult. Of the three teachers who did, one explained it was difficult because he "wasn't used to it."

At this point of admission to student teaching, we wondered if the assessment process produced any information for the teachers regarding their strengths and weaknesses as a prospective teacher. Almost all of the teachers said, "no." One teacher who had received such information said it was provided by a supervisor who had videotaped her teaching two lessons during the summer when she worked as a teaching assistant. She said she found the information to be very useful because "solo teaching was new to me," and also "very enlightening" because the videotape, along with the discussion with her supervisor, gave her a sense of her classroom presence and her communication skills. Two other teachers said they received information about their strengths and weaknesses from an instructor or professor, and that the information was determined from their presentations of lessons in class and other interactive classwork. Both teachers described the information they received as accurate, fair and useful.

Student Teaching

Before interviewing the teachers about the assessments they experienced during student teaching, we asked them to give us a little contextual information about each of their student teaching assignments. This information is presented in Table 5.2, in a form akin to a Harper's Index. Three pieces of information stand out from this data: (1) the typical amount of time spent at the school site during student teaching varies significantly between elementary teachers and secondary teachers, with the former spending a minimum of a half day at the school and the latter only two periods, (2) the majority of master teachers did not have much supervisory experience (i.e., five years or less), and (3) almost one third of the master teachers were supervising a student teacher for the first time.

TABLE 5.2

STUDENT TEACHING: FACTS AND FIGURES

(Based on Data from Teacher Interviews)

Average number of student teaching assignments per teacher: 2

Range of assignment length: 6 weeks to 9 months (i.e., one school year)

Typical amount of time spent at school site: Elementary teachers – 1/2 to full day
Secondary teachers – 2–4 periods a day

Typical student population: Ethnic mix of low to middle class students, remedial and/or regular students

Typical university supervisor: University faculty member

Number of years supervisory experience: 50% of teachers said their supervisors had six years or less; 25% of teachers had no idea

Typical amount of contact with university supervisor prior to student teaching: None

Percentage of teachers who said their university supervisor(s) had same subject-matter background as they did: A little over 50%

The subject area in which there were the fewest cases of match-up: History/social sciences

Total number of master teachers supervising the sample teachers: 44

Number of years teaching experience: 65% had 10+ years

Number of years supervisory experience: 65% had five years or less

Percentage serving as master teachers for first time: 33%

Typical way student teacher is assigned to master teacher: University assignment

Percentage of teachers who chose a master teacher: 42%

Percentage of teachers who said they had no contact with their master teachers prior to student teaching: 70%

The first point of information is confirmed by the data collected in the IHE study. The second two points were not confirmed for accuracy, but may reflect the difficulties in recruiting a pool of master teachers that were described in the IHE study. Even assuming that some of the teachers were mistaken in their judgements of their master teachers' supervisory experience, the information indicates that a fairly large number of master teachers were charged with assessing student teachers' knowledge, skills and/or abilities even though they had limited experience in this role. More information about the master teachers and their role in assessment is provided below, following a discussion of assessment by university supervisors. Also included in this section is a discussion, based on the teachers' recollections and opinions, of the teachers' overall experiences with assessment during student teaching.

Assessment by university supervisors. During student teaching, every teacher was assessed by someone affiliated with the teacher education program, a person whom we refer to as a university supervisor. Prior to beginning student teaching, most teachers received information as to how they would be assessed by their university supervisor(s), often through a student teaching handbook or an orientation meeting. One teacher said she knew of the method her university supervisor would use to assess her because it was the same method her instructors used to assess her in her education courses. Most teachers also received information about the assessment criteria that their university supervisors would be using, sometimes by being given a copy of the evaluation form that their supervisors would be using.

All of the university supervisors used the observation method to assess the student teachers, but the number of observations made by each university supervisor differed greatly. The range seemed to be two observations in eighteen weeks (experienced by a secondary science teacher) to approximately one observation every week (experienced by several teachers). For secondary teachers, the number of observations also sometimes varied by class. One teacher, for example, who student taught in two different high school science classes (i.e., biology and human biology), said his university supervisor observed him three times in the class with the "lower-end" students, but did not observe him at all in his other class. Most observations lasted one period (approximately 50-55 minutes), but many teachers cited observations lasting less than one period, i.e., 15 to 30 minutes. Shorter observations were usually of the informal kind, but not always. No teacher recalled an observation by their university supervisor lasting more than one hour.

While the primary method of assessment was observation, the majority of teachers said that their supervisors assessed their teaching skills using other methods of assessment

as well. Methods named included examination of lesson plans, talking to master teachers and the students, requiring and reading logs, and, according to one teacher, taking Polaroid pictures. Assessment by videotape was also experienced by several teachers, although one teacher said his supervisor only videotaped those teachers having "control problems" (which he didn't have). Another teacher described how a videotape, which was made by the university supervisor as she simultaneously conducted the observation, was used in an assessment of her teaching skills:

> *[I] had to watch the videotape three times and write a 7-8 page essay analyzing the lesson: what I did right, what I did wrong, what might I change next time.*

She described this method of assessment as "extremely helpful," but added that "we all hated it." However, she also stated that the videotape assessment is what she remembers most about student teaching in terms of helping her in her current job.

The teachers varied in how they valued their university supervisors' assessment of their teaching skills. Now teaching in classrooms of their own, the teachers were asked to reflect on how their supervisor's assessment of their teaching skills has been most helpful to them. Most teachers cited at least one specific area--and sometimes two or three--in which the supervisors' assessment had been useful. The area most frequently mentioned was that of pedagogy, especially classroom management. Several teachers also said their university supervisors' assessment helped them in working with students, and other teachers commented that their supervisors' assessment of their teaching skills helped them the most by giving them confidence in themselves as teachers. A few teachers remarked that their university supervisors' assessment helped them to be more reflective practitioners, either with regard to the lessons they taught or with regard to their general teaching style and the way they presented themselves to students.

Approximately 25% of the teachers saw no value at all in their university supervisors' assessment of their teaching skills. Commented one of these teachers, "I divorced myself from that whole thing and started anew."

Even many of the teachers who had benefitted from their supervisors' assessment of their teaching skills identified areas in which their supervisors' assessment was not especially helpful. The two areas mentioned most often were (1) content pedagogy (or in the words of the teachers, "subject matter help"), and (2) working with diverse students. In

the first area, teachers, both elementary and secondary, stated that they "would have liked more specific suggestions re: dealing with specific subject areas." For example, three of the five teachers who student taught in secondary social studies classes faulted their university supervisors for not providing much subject matter help. Not surprisingly, all three teachers said their supervisors were lacking social science backgrounds.

Working with diverse students was an area especially noted by two of the three teachers who completed their student teaching in out-of-state teacher education programs and who are now teaching in California regions with very diverse student populations. Both teachers commented that they wished their university supervisors had provided them with more information about working with students of different ethnic groups. One of the three teachers who student taught in bilingual classes also saw a need for improvement in this area. This teacher said his supervisor's assessment of his teaching skills was deficient in information about working with at-risk students and bilingual materials--information he sorely needed as a student teacher in an inner city middle school. As a result, this teacher claimed that there was nothing about his supervisor's assessment of his teaching skills that he found to be the least bit helpful. (Of the other two bilingual teachers, one said her university supervisor's assessment of her teaching skills had "no influence" on her at all, and the other commented that the most helpful part of her supervisor's assessment was that it "encouraged me to stick with young students.")

Other teachers stated that their university supervisors' assessments would have been more helpful if they had been less general and had provided more specific suggestions, ideas, and criticism. For example, one high school teacher, who had some serious classroom management problems during his student teaching, observed, "The evaluations weren't that helpful because they were more theoretically oriented than practical." Finally, two teachers faulted their supervisors' assessments of their teaching skills because they did not believe their supervisors visited their classes often enough to accurately judge what was happening on a "day-to-day basis," and one teacher said his supervisor was least helpful in the area of providing information about how to evaluate students.

Assessment by master teachers. Except for the intern teacher, all of the teachers in the sample were assessed by at least one master teacher. In contrast to the many teachers who said they were informed prior to student teaching about how their university supervisors would assess them, only 50% of the teachers said they knew how their master teacher(s) were going to assess them. One teacher commented, "I'm not sure even now how the process works ideally."

Of those teachers who were informed about how their master teachers were going to assess them, most said they were also aware of the assessment criteria that their master teachers would be using, sometimes because they received the evaluation form that the master teacher was expected to use.

Unlike the university supervisors who observed the teachers once a week at best, the master teachers were in the position of being able to observe the teachers and talk with them about their teaching skills on a daily basis. In fact, 60% of the teachers specifically stated that at least one of their master teachers spoke with them about their teaching skills every day. This frequent communication was especially appreciated by some of the teachers who described the rate as "excellent" and described their master teachers as "on top of things."

The rest of the teachers said their master teachers spoke with them about their teaching skills either several times a week or once a week. Approximately 25% of the teachers also reported that the frequency of communication changed over the duration of the student teaching period, usually being more frequent at the beginning and decreasing towards the end when, as a couple of teachers explained, you "get better" and "need less." One teacher, however, described as "disappointing" the fact that her master teacher "gave me feedback for a while, and then just disappeared and left me alone so she could work on the school sports program." (This teacher also related that, while absent, her master teacher was secretly tape recording her lessons--a discovery accidentally made by the teacher when she went to play a tape for her students!)

The teachers valued their master teachers' assessment of their teaching skills for a variety of reasons. Although some of these reasons were identical to the ones given for valuing their university supervisors' assessments--i.e., pedagogical help (the response most frequently given), help with working with students, and help in boosting their confidence in themselves as teachers--other reasons were markedly different. Approximately 30% of the teachers, for example, valued their master teachers' assessments because of the specific subject-matter help they provided, as the following comments indicate:

> *Showed me different ways to approach subject matter.*
>
> *I learned a lot about teaching reading and hands-on activities in science and math.*
>
> *He exposed me to a lot of resources in subject matter area.*

Taking all this information into consideration, it is not surprising that almost 75% of the teachers said that their master teachers' assessment of their teaching skills was more useful to them than their university supervisors'. Perhaps reflecting the fact that university supervisors are not in the classroom as often as master teachers and hence do not have an equivalent role, the teachers' reasons for valuing their master teachers' assessments were as follows: 1) because the master teachers saw the student teachers on a daily basis, the student teachers believed they knew more about their teaching than did the university supervisors (who were more likely, according to the teachers, to see a "staged lesson"), 2) the master teachers were perceived as being more in touch with the classroom, the students, and the situation, 3) the master teachers gave more practical, specific, and frequent suggestions, and 4) the master teachers modelled specific techniques and behaviors. Some of the teachers' comments explaining why they preferred their master teachers' assessments are as follows:

> *On a daily basis, I saw her teach and respected her opinion. I didn't see my university supervisor teach.*
>
> *My master teacher was there more--knew what teaching was, what problems were, what was more important.*
>
> *My master teacher had a more practically-based approach.*

Interestingly, those teachers who found their university supervisors' assessments to be more useful generally did so because they said their supervisors gave more specific and practical information than did their master teachers. One teacher, however, preferred her university supervisor's assessment for a completely different reason: she did not like her master teacher's assessment of her teaching skills and considered her university supervisor's to be more "unbiased."

This teacher's dismissal of her master teacher's assessment simply because she didn't agree with it is a good example of how very often the teachers did not hold their master teachers' or university supervisors' opinions in high regard and, thus, could easily ignore them. Another good example is the first quote above where the teacher's respect for the master teacher's opinion is a result of having observed the master teacher's teaching skills and liking what was seen. This teacher had not seen the university supervisor teach and thus respect was not automatically accorded. This sentiment that university supervisors, in particular, had to earn respect seemed quite prevalent. One teacher who mentioned some positive things he gained from his university supervisor's assessment still

stated that, overall, he didn't think much of her skills and, as a result, "I took her advice with a grain of salt."

In general, the teachers did not seem to recognize the master teachers and university supervisors as "experts" whose advice was to be trusted and valued. There may be several reasons for this. One very likely reason may be reality: Some master teachers and university supervisors are in their positions not because they are experts but because someone had to fill the role. Another possible reason, however, is that in the teaching profession there is a lack of definition of what constitutes an expert. Few people automatically equate longevity in the teaching profession with expertise. For some people, an expert is defined by their classroom management; for others, an expert is defined by their students' creativity.

Still another possible reason is that in teaching there is a lack of a framework or common language with which to discuss teaching skills. For example, the teachers rarely described the feedback they received in terms of concepts (with the sole exception of classroom management, which was a term used by most). Instead, they usually described the feedback in terms of specific suggestions or techniques, e.g., how to make a presentation more interesting or how to work with LEP students. If the way they described feedback is at all reflective of the way they think about teaching, then teachers seem to view teaching as a hodgepodge of disparate skills. That is, they seem to lack a framework through which to view teaching in general, and their teaching in particular. Such a lack of framework might easily facilitate the rejection of negative feedback about specifics because a teacher is unable to see the importance of the feedback in relationship to a larger group of skills. Thus, beginning teachers very often have to apply their own interpretation as to what the feedback from their university supervisors and master teachers means (i.e., its overall significance), and their own judgement as to whether to trust the feedback or not.

Overall assessment experience. After interviewing the teachers about their university supervisors and master teachers, we asked them to reflect on the results of the assessments made in their first and last (usually their second) student teaching assignments and to comment on their overall experience of being assessed during those assignments. Almost all of the teachers said that the evaluation of their strengths and weaknesses made at the end of their first student teaching assignment was accurate, complete, and fair. This positive reaction may be in part due to the fact that the most commonly recalled strength was "good student rapport" -- a characteristic that few, if any, teachers would likely dispute as inaccurate, incomplete or unfair. Of the few teachers who did disagree with their evaluations, one said he "didn't get a sense of what I needed to work on from them" and two

faulted the evaluations they received because they felt they weren't complete (i.e., a math teacher said his university supervisor didn't know enough math to make a complete assessment, and an elementary teacher judged her assessment incomplete because her master teacher "wouldn't let go, so she didn't do much teaching").

When asked if they considered the assessment made at the end of their first student teaching assignment to be rigorous and demanding or weak and unchallenging, only two of the teachers responded "weak and unchallenging" (see Table 5.3). (It should be mentioned that we did not define the terms, "rigorous," "weak," etc., so that the teachers' interpretations of the words were likely to have varied.) The remaining teachers were split as to whether it was rigorous or somewhere between rigorous and weak. One teacher explained why he was undecided:

> *The hurdles you had to jump were unclear. The experience was challenging, but the evaluation wasn't. The university supervisor and master teacher had different standards so it made it more difficult. The master teacher's was less challenging, the university supervisor's more so, but I was never sure about what was most important because there was no grand overview of teaching.*

Perhaps because so many teachers perceived their evaluations as something less than rigorous and demanding, many of the teachers offered suggestions for improving the way they were assessed during their first student teaching assignment. For the most part, these suggestions echoed the ones described earlier with regard to the teachers' suggestions for improving their university supervisor's and/or master teacher's assessment of their teaching skills: increase the number of observations, increase the number of people who make the observations (e.g., have at least one other person besides the university supervisor make an observation), and give more specific feedback.

Two teachers said they would have benefitted from being videotaped either routinely or at least at the beginning and end of the assignment, and two other teachers said they would have benefitted from more responsibility, i.e., more teaching time.

Not all of the teachers had a second student teaching assignment (e.g., some teachers' assignments lasted the school year), but the majority of teachers did, and almost all of these teachers said the second evaluation of their strengths and weaknesses was different from the first. (Note: A few secondary teachers considered their second class, if a different subject from their first, to be their second student teaching assignment.) The

TABLE 5.3

TEACHERS' PERCEPTIONS OF THE RIGOR OF ASSESSMENTS

Point of Assessment	Number of Teachers Who Said: "Assessment was Rigorous and Demanding"	"In-Between"	"Assessment was Weak and Unchallenging"	"Undecided" or N/A
Admission to Teacher Education Program (N=21)	7 (33%)	2 (10%)	10 (47%)	2 (10%)
First Student Teaching Assignment (N=21)	10 (47%)	8 (38%)	2 (10%)	1 (5%)
Second Student Teaching Assignment (N=21)	10 (47%)	5 (24%)	1 (5%)	5 (24%)
First Year of Teaching (N=22)	5 (23%)	5 (23%)	9 (41%)	3 (14%)

differences noted tended to fall into three categories. First, the difference most frequently mentioned was that the teachers were perceived as having greater or more abilities after their second student teaching assignment. For example, one teacher noted,

> *My second evaluation said that I had developed student-centered activities and used them. I learned to teach in methods other than lecture. And I used more cooperative learning.*

A second type of difference was that the scores from the second evaluation were either higher or lower than the scores on the first. And a third type of difference was that the weaknesses and/or strengths described after the first assignment were not the same as those described after the second. Some sample comments:

> *My strengths were the same, but the weaknesses were different. Flexibility was not mentioned in the second evaluation; instead, I moved students too fast.*

> *Classroom management was weak in one class, but fine in the other. I also think that student rapport was lower than after my first class.*

As evidenced by these last two examples, the quality of a teacher's performance could and sometimes did change depending on the class being taught. Although none of the teachers expressed confusion as a result of any of the differences noted, the potential for confusion clearly seems to be there, especially if the second evaluation is less positive than the first. For example, if a teacher's classroom management skills are evaluated more positively after the second assignment, the conclusion usually is made that the teacher's skills have improved. But what if a teacher's classroom management skills are evaluated more negatively after the second assignment? Does this mean the teacher's skills have deteriorated from the first assignment to the second? What conclusion is the teacher to make from two evaluations which are totally different?

It seems that if the teachers are to make sense of their assessments, there needs to be a framework against which they can interpret their feedback. Classroom management, for example, could be presented as a skill that is developed over time, as well as a skill which looks different in different situations. The skill of classroom management could be defined in progressive terms so that what might be expected of a teacher during the first student teaching assignment is different from what is expected in the second. With such a framework of assessment, teachers might be better able to understand the significance of two very different evaluations of the same construct.

Despite the noted differences, most of the teachers still considered the evaluation made of their strengths and weaknesses after their second student teaching assignment to be accurate, complete, and fair. Of the few teachers who disagreed, the remarks most commonly made were either that the assessment was not complete (e.g., "she could point out more weaknesses than she did") or not fair (e.g., "the end evaluation didn't correspond to daily feedback").

As was the case with the first student teaching assignment, the teachers were asked if they considered the evaluation of their strengths and weaknesses after their second student teaching assignment to be rigorous and demanding or weak and unchallenging. As Table 5.3 indicates, the majority of teachers who responded (i.e., had a second assignment) considered the assessment made after their second assignment to be rigorous and demanding. Although, as mentioned earlier, the terms, "rigorous," "weak," etc. were open to teacher interpretation, some of the teachers' comments about the assessment made after their second student teaching assignment give an indication of what constitutes a rigorous assessment. For example, for some teachers, the rigor was a result of the number of areas that were assessed:

> *The second was pretty challenging. There were many more categories. I was unaware of 20 of them to say the least. There were things that I hadn't thought of.*

For other teachers, the rigor was a result of the amount of work expected:

> *The second was rigorous. Really a lot of work. I had to turn in so many lesson plans, and I had to follow a format.*

A few teachers also described as rigorous the expectations that were held of them:

> *The second was rigorous because a lot was expected. I had to perform at a high level to get a good evaluation.*

Only one teacher described the assessment as weak and unchallenging, and, compared to the first assessment, a smaller number of teachers described the second assessment as "in between" rigorous and demanding.

The suggestions offered by the teachers for improving the way they were assessed during their second student teaching assignment were, not surprisingly, almost exactly the

same as those made for the first assignment: more observations, more feedback, longer observations by the supervisor, and more observations by other staff members who are familiar with students and the situation.

One suggestion, however, stood out from the rest. This suggestion was made by a teacher whose master teacher became a vice-principal two weeks after the start of her second assignment and was never replaced, leaving her in sole charge of the classroom. Her suggestion: "I would have liked to have had a master teacher." Unfortunately, this teacher wasn't the only one who lacked the benefit of sufficient or proper supervision. There were others: a teacher whose master teacher decided to spend more time with the school's track team than observing the teacher in her second assignment; a teacher whose university supervisor only observed in one of his classes and never in the other; and a teacher who was told that, for her second assignment, she was being assigned to a master teacher with lots of experience but who needed help in dealing with problem students in the class. Although these are only four examples, four out of 21 teachers is a fairly high percentage if it is at all representative of the percentage of teachers across the state who may have experienced a lack of sufficient or proper supervision. It is one thing for a student teacher to experience some observation and supervision, and then to desire more, and quite another thing for a student teacher to experience a lack of supervision and to desire some.

Employment Assessments

Upon completion of their teacher education program--or sometimes just prior to completion--the teachers began applying to school districts for a teaching position. A little more than half of the teachers applied to five districts or less and usually to districts in the same geographical area where they were residing. The big exception to this were the teachers who attended out-of-state programs. Two of the three teachers from out-of-state said they applied to between 30 and 40 districts in and out of their home state. Some of these applications were completed on campus during job recruitment fairs which personnel representatives from numerous states attended.

When asked to describe the process by which they were hired for their first regular teaching job, almost all of the teachers said that they were interviewed for the job by a school administrator (e.g., principal, vice-principal) and/or district personnel (e.g., superintendent, personnel director, director of bilingual education). Slightly less than half of the teachers said that their interview process included interviews with other teachers. In fact, less than half of the secondary teachers (5 of 13) recalled being interviewed by a teacher with the same subject matter background.

Aside from the interview process, assessment of the teachers' competency seemed to primarily rely on the letters of reference required by most of the districts. Only six of the teachers, for example, experienced a demonstration or performance assessment. Three of these teachers were observed by their future principal while they substituted or student taught, two were specifically requested to give a demonstration lesson in a class with which they were unfamiliar, and one, who was applying for a bilingual position, was asked to demonstrate her Spanish skills by giving directions to a child in Spanish.

Once the teachers were hired, we wondered if they received any information about their strengths and weaknesses as identified during the hiring process. The overwhelming answer was "no." Of the three teachers who said they received information, none mentioned any weaknesses, and only one named some specific strengths: "He picked me because he liked the way I gave directions and my sense of humor." (This teacher was the bilingual teacher described above who was asked to give directions in Spanish as part of her interview process.)

First-year Assessments

Having been hired, the teachers settled in to begin their first year of teaching--and their first year of assessment by a school administrator. Before we describe the assessments they experienced, however, we will briefly describe some of the diverse contexts in which the teachers were teaching. At the elementary level, the context ranged from a rural second-grade class of 14 students described by the teacher as mostly "upper middle class to very wealthy Caucasians," to a fifth-grade Spanish bilingual class in the Los Angeles area full of students coming from the barrio whom the teacher described as having "serious behavior problems" and "very poor." At the middle school level, the diversity in context included a rural school where the teacher taught the school's sixth, seventh, and eighth graders "everything but math and science", and a suburban magnet school devoted to computers, science and math, and composed of 50% GATE students, in which the teacher only taught 8th-grade English. At the high school level, the contexts were equally diverse, ranging from a social studies teacher teaching in an inner city school full of "gang-oriented" students, many of whom read at a fifth-grade level or lower, to an English teacher teaching in a predominantly Caucasian suburban school populated by "regular" students from the "high end of middle class."

Despite the wide differences in teaching context, there were few differences in the way the teachers were assessed during their first year of teaching. For example, except for one rural school teacher who said he was never formally evaluated (he said his community

was small enough so that "people know whether or not you can teach"), the majority of teachers were formally assessed through observation two to three times in the year by their principal or vice-principal. Most of these observations, like those conducted during student teaching by their university supervisors, lasted from 40 to 55 minutes, although close to 40% of the teachers said they were observed for 30 minutes or less. The majority of teachers also said their assessments included a post-observation conference, with almost 40% reporting having a pre-observation conference as well.

The most notable difference in the assessment process experienced by the teachers seemed to be in the time of year when their first assessment was conducted. Although the majority of teachers experienced their first observation during the first quarter of school, approximately 20% of the teachers said their first observation occurred in the first month of school, and approximately 40% said their first observation did not take place until November. Three teachers reported being observed for the first time in January or later.

One other difference worth noting: Although most teachers were observed by their principal or vice-principal, a few teachers mentioned being observed by someone different, usually someone with a similar subject matter background. One teacher, for example, said her department chair conducted six informal observations of her, each lasting 10 to 60 minutes. Another teacher, whose first year was a difficult one, said he was observed 10 to 11 times during the first year, and that his curriculum leader conducted some of the observations. Sometimes the someone different was an outsider as was the case with still another teacher who reported being formally observed five times during her first year by a CNTP teacher consultant who did not work in her school. This consultant also made "numerous informal observations" and corresponded with the teacher through a journal (e.g., the consultant sometimes answered questions in the journal instead of conducting a post-conference with the teacher).

When asked what feedback they were given about their strengths and weaknesses after their first observation, the teachers gave various responses. Most of the strengths identified revolved around how the teacher related to students or involved the students in activities (e.g., "good cooperative groups," "kept kids on task"). The most common weaknesses identified were linked to classroom management (i.e., "how I managed a troublemaker during the lesson") and execution of the lesson observed (e.g., "lack of closure"). A little more than 25% (6 of 22) of the teachers said they had no weaknesses identified. Interestingly, if we compare the strengths and weaknesses that were identified after the teachers' first evaluation as a teacher with those that were identified after their first student teaching assignment, a little over 40% (9 of 21) of the teachers reported having

a same strength identified in both assessments, while two teachers had a same weakness identified. The strength usually identified was a rapport with students or an ability to involve students in activities, and two teachers were recognized in both assessments for their subject matter knowledge. One teacher was identified by both assessments as being weak in discipline, and the other teacher's weakness was classroom presence, e.g., nervousness, stiffness. Although the similarity in strengths is positive in that consistency is implied, it may be negative in that some duplication of assessment is also implied.

Close to 60% (13 of 22) of the teachers said the information they received from their first observation was helpful. Only two of the teachers, however, said they benefitted from the information because they learned new things to help them with instruction (e.g., "the principal had good points about closure," "it was good to know I called on that girl too often"). Instead, the majority of teachers found the information helpful because it gave them confidence and a feeling of being supported. Of those teachers who did not find the information helpful, most faulted the information for being too general (e.g., "it was just a generic checklist," "it wasn't specific").

Most of the teachers who reported some identified weakness(es) also said they were offered sources of help. Most frequently, the teachers said that the same person who identified the weakness (e.g., the principal or vice-principal) made recommendations on how to address it. Sometimes, however, help came from other teachers or district resource people. A slight majority of all the teachers (59%), for example, said they received guidance and assistance from a mentor teacher or other experienced person during their first year. (Over half of these teachers were participants in a California New Teacher Project). Two teachers with identified weaknesses said they were told to observe other teachers, but only one of the teachers said coverage was provided to do so.

After their final evaluation during the first year, a little over 40% of the teachers (9 of 21) said they had no strengths identified or that they couldn't recall any identified strengths. Several of these teachers did recall "stuff being checked off on the Stull bill," but they didn't recall what the "stuff" was, only that "everything was marked OK." Rapport with students (or working with "this type of students") was again the most commonly identified strength, and enthusiasm came in a close second. Three of the 21 teachers described strengths pertaining to their instructional skills--"lectured well," "good cooperative groups," and "well organized labs"--and two teachers were commended for their subject matter knowledge.

The majority of teachers said there were no weaknesses identified after their final evaluation. Of the six teachers who did report identified weaknesses, four pertained to classroom management or discipline. One other teacher was faulted for "lecturing so much" and another for having trouble "managing paperwork." Only one of the teachers said the weakness identified after the final evaluation was the same as the one identified after the first evaluation.

Overall, looking at the strengths and weaknesses identified after the first observation and after the final observation, there seems no discernible pattern to the constructs assessed. Furthermore, based on the teachers' responses, there is nothing to indicate that either the teachers or their assessors consider teaching skills to be of a developmental nature. Indeed, the lack in number and variety of weaknesses identified after the final evaluation seems to suggest just the opposite.

Despite the rather general nature of the final evaluation, slightly more teachers described the information from the evaluation to be helpful than did not. Again, however, the help was more often a feeling of support and encouragement and an increase in confidence rather than specific ways to improve their teaching.

How did the teachers judge the quality of their first year assessments? Did they see them as rigorous and demanding? Or weak and unchallenging? Looking again at Table 5.3, we see that the response most frequently given was to describe the assessments as "weak and unchallenging." (See Table 5.4 for a comparison of teachers' perceptions of the rigor of assessments across assessments.)

Some of the teachers who gave this response explained their answers:

> *He just came in and watched me. I'd like to have been watched by others and watched others or had to keep a portfolio. I could bluff 3-4x a year; it's easy to do good when he only comes in 3 or 4 times.*
>
> *I'm sure the vice-principal has had experience as a teacher, but not for 10 years. I would have preferred to have been evaluated by a practicing teacher familiar with current practice.*
>
> *Not enough feedback. Evaluations are cursory. There weren't enough suggestions for improvement.*

TABLE 5.4

TEACHERS' PERCEPTIONS OF THE RIGOR OF ASSESSMENTS ACROSS ASSESSMENTS

Teacher	Admission to T. Ed. Program	First Student Teaching Assignment	Second Student Teaching Assignment	First Year Assessments
1	W	I	NA	W
2	W	R	NA	W
3	DK	R	I	I
4	W	I	I	W
5	W	I	NA	W
6	R	R	W	W
7	R	R	NA	W
8	W	I	R	I
9	DK	I	I	W
10	W	R	R	W
11	I	I	R	I
12	W	W	NA	NA
13	R	W	R	R
14	W	R	I	I
15	I	R	R	R
16	R	R	R	R
17	R	R	R	DK
18	W	DK	R	R
19	R	I	I	NA
20 (intern)	NA	NA	NA	W
21	W	I	R	I
22	R	R	R	R

W = "weak or unchallenging" R = "rigorous and demanding"
I = "in-between" DK = "don't know"
NA = not applicable

Of the remaining teachers, an equal number described their assessments as rigorous and demanding or "in between." (One teacher could not decide.) A secondary teacher who was observed six times during her first year by a CNTP teacher consultant explained why she described her assessment as rigorous:

> *Because I was observed so frequently and she always looked for an area of improvement. Nothing was ever left to stand still. They demand excellence of teachers and teachers demand it of themselves.*

When asked if there were any improvements they would make in the way they were assessed during their first year of teaching, all but four of the teachers responded affirmatively. Altogether a total of eight different suggestions for improvement were made. Each suggestion is listed below along with the number of teachers who made them:

Suggestions for Improving First-Year Assessments

More detailed feedback	(9)
More observations	(6)
Observations conducted by different people	(5)
Longer conferences	(1)
Longer observations	(1)
Fewer formal observations; more informal observations	(1)
Assessor with same subject-matter background	(1)
First observation before end of semester	(1)

What is immediately obvious, is that these suggestions are almost identical to the ones made earlier for improving the way assessments are made during student teaching. As the table indicates, half of the 18 teachers who offered suggestions of improvement,

expressed a desire for more detailed feedback after the observations. Some teachers expressed a desire for more specific feedback in general, such as one elementary teacher who was observed only once: "I would have liked to have had specific constructive advice so I knew what to work on and how." Other teachers were interested in receiving detailed feedback in specific areas, such as a high school teacher who was observed several times by her vice-principal:

> *I would have liked more details of performance; a more detailed, outlined view of classroom relations to make sure I was involving all students as opposed to being the teacher behind the podium. I would have liked more specific feedback on this.*

Of the teachers who wanted more observations, half were those who had been observed less than three times, and the other half had experienced more than three. One middle school teacher who was only observed twice thought that more observations would have "made me more comfortable when someone is observing" and would have "helped build up my confidence, too." A high school teacher who was observed three times by his vice-principal said he would have liked to have been observed "every month on average." He particularly would have liked to have been observed teaching a series of lessons over a three-day period and then to have received specific feedback on how he did.

The teachers who wanted observations conducted by different people usually meant that they would have liked at least one other person, such as a mentor or peer teacher, to observe their teaching. Commented one middle school teacher assessed three times by his principal:

> *I would have different people assess me. I think the principal liked me. I would have liked to have been observed by additional people.*

On the other hand, two teachers who were observed once by their principal and then five or six times by another person (i.e., a department chair; a teacher consultant), expressed a desire for more observations by their principal. Overall, the general feeling seemed to be that teachers would gain more if they did not have to primarily rely on one person for feedback. (This viewpoint was also shared by one of the four teachers who were satisfied with the way they were assessed during their first year. This teacher, a science resource teacher in an elementary school who went from class to class conducting his lessons, was only observed twice by his principal, but said he "wouldn't change a thing" about the way he

was assessed because "I felt like I was evaluated by the whole staff eventually," and he received "fantastic help" from them.

Finally, after having talked with the teachers about their assessment experiences, both during student teaching and in their first year of teaching, we asked them to comment on which assessments were most helpful to them and why. The majority of teachers named the assessments they experienced during student teaching as the most helpful, and almost all gave the following reason (or a variation thereof): in student teaching, one receives more specific and frequent feedback about one's teaching skills. Whether the feedback was from a master teacher or a university supervisor didn't matter; what was of primary importance was that feedback was frequent and specific, as these comments indicate:

> *Student teaching was the most helpful because my university supervisor made frequent observations and gave relevant feedback.*

> *My master teacher's assessment in my second student teaching assignment was most helpful. We communicated daily and he gave me specific feedback and help with materials. I could use all that information in my own class.*

Of those teachers (approximately 20%) who said their first-year assessment experiences were the most helpful, various reasons were given. The most common was that the principal's assessment had direct influence on their job and thus it was helpful to the teachers to know what they had to do to keep their job. One elementary teacher, for example, said that she took "to heart" what her principal said the first year because it was "applicable to the grade level and position" in which she wanted to remain. Somewhat similarly, another elementary teacher said he found his principal's first-year assessments to be more helpful than those he experienced during student teaching because he considered student teaching to be a "totally unrealistic experience."

Another 20% of the teachers said their student teaching assessments and those experienced during their first year of teaching were equally helpful. Commented one middle school teacher:

> *It was helpful to have people observe me and talk about what was going on both in student teaching and the first year.*

This teacher's comment echoes the sentiment expressed by so many of the teachers that observation assessments, whether conducted by university supervisors, master teachers, or school administrators, are desired as long as they are accompanied by feedback.

Assessment Content

In our interviews with the teachers, we attempted to learn about the content of the assessments they experienced by asking them specific questions about the content of the feedback they received at certain assessment points. In particular, we focused on the points of assessment beginning with student teaching through their first year of teaching. The questions that were asked either focused on the assessment feedback they might have received in specific areas (e.g., general pedagogy, content pedagogy, working with diverse students) or requested general descriptions of any feedback they received from the assessments about their strengths and weaknesses. Although the teachers seemed to try their best to answer these questions, it sometimes seemed very difficult for them. In particular, there seemed to be a lack of a framework or language in which to discuss the content of teaching. Nevertheless, their answers did provide enough information for some discussion in three content areas: (1) general pedagogy, (2) content pedagogy, and (3) working with diverse students. In addition, we also present a discussion of how the content of the state curriculum frameworks is or is not included, **according to the teachers**, in the content of the assessments.

General Pedagogy

The first time we focused the teachers' attention on general pedagogy was when we interviewed them about the assessments they experienced during student teaching. While close to 70% of the teachers said that their supervisors focused more on assessing their general pedagogical skills than on any other skill (the teachers were given several skill areas to choose from), the majority of teachers seemed relatively inarticulate when it came to describing those skills. For example, in the interview, we asked the teachers to describe any feedback they received from their university supervisors about the development of their general pedagogical skills. Not sure if they would fully understand the term "pedagogy," we loosely defined the term as "basic teaching skills" and gave such examples as classroom management, rapport with students, the establishment of routines, and developing lesson plans. The most frequent response given by the teachers was an echo of our first example: classroom management. Only four of the 21 teachers described any type of pedagogical feedback other than classroom management or discipline. These four teachers said their university supervisors gave them feedback about such pedagogical skills as "running groups,"

"the teacher's movement in the class," "routines," "number of students on task," "rapport with students," and "instructional techniques." A few teachers mentioned that they got "a lot" of pedagogical feedback in written form or in post-conferences, but they did not specify the content of the feedback. Two teachers said they did not receive, or did not recall receiving, any feedback at all from their university supervisors in the area of pedagogy.

The teachers responded in almost exactly the same way when asked to describe the general pedagogical feedback they received from their master teachers. Although the teachers reported that their master teachers were less unidimensional with regard to the focus of their assessments than were their supervisors, about a third of the master teachers were still reported by the teachers to have focused primarily on assessing their general pedagogical skills and another third to have focused primarily on general pedagogical skills and one other skill area (i.e., teaching a particular subject or working with diverse students). Despite this emphasis on general pedagogical skills, however, the teachers were again relatively inarticulate when it came to describing the general pedagogical feedback they received. The response most commonly given by the teachers when asked to describe the type of general pedagogical feedback they had received from their master teachers was "classroom management." Several teachers also responded to the question by saying that they received "a lot" of general pedagogical feedback from their master teachers, and when asked to elaborate, specified the areas of classroom management or routines. Only three of the 22 teachers identified more than two areas in which they received general pedagogical feedback from their master teachers (e.g., "cooperative learning, discipline, grading"; "routines, rapport with students, monitoring students").

According to the teachers, classroom management was also the major pedagogical focus of the questions they were asked when interviewed for their first job. When we asked the teachers to describe the types of questions they were asked in their hiring interviews about their general pedagogical skills, the majority of teachers said they were either asked to describe their "classroom management style" or "classroom management system," or to respond to "situational questions" such as "If this happened, what would you do?" Only five teachers recalled being asked about any other general pedagogical areas other than classroom management. Some of the general pedagogical areas they identified were: "types of teaching tools (group activities, lecture styles)," "homework," "learning/teaching strategies," "rapport with students," and "lesson plans."

In our conversations with the teachers about the assessments they experienced during their first year of teaching, we asked them to talk about any information they received from the assessments regarding their strengths and weaknesses as a teacher. In

particular, we asked them to describe their strengths and weaknesses as they were identified after their first assessment of the year and after their last. Although many of the teachers said they received no information at all about their strengths and/or weaknesses either after their first assessment or after their last (especially the latter), of those teachers who did receive information, most described strengths or weaknesses that were pedagogical in nature. For example, after their first assessment some of the teachers were told that their strengths and weaknesses were as follows:

Strengths	Weaknesses
kept kids on task	went too fast in lesson
well-organized lesson	lack of closure
cooperative learning	classroom management
ability to control kids	discipline
good classroom rapport	questioning techniques

Although each of these examples represents a particular pedagogical area -- i.e., planning, management, delivering instruction, or evaluation -- few of the teachers reported strengths and/or weaknesses in more than one area.

In summation, from student teaching to the first year of teaching, general pedagogy was the major focus of all the assessments that the teachers experienced. Moreover, from the teachers' perspective, classroom management was the area of pedagogy most often assessed and remarked upon by the assessors. Although examples from other areas of pedagogy such as planning, delivering instruction, and evaluation were also mentioned, the majority of teachers seemed to recall classroom management as the primary focus of their assessments, especially those made during student teaching and when they were hired. In their first year of teaching, pedagogy in general seemed to be the focus of assessment rather than classroom management in particular.

Content Pedagogy

For the purposes of our study, content pedagogy meant skills in teaching a particular subject or subjects. The questions we asked the teachers about content pedagogy were the same as those asked about pedagogy in general. That is, at the different assessment points beginning with student teaching, we asked the teachers to describe the feedback they received or the questions they were asked about their skills in teaching specific subjects.

Content pedagogy as a focus of assessment was experienced much less often by the teachers than was general pedagogy. In fact, only two teachers said their university supervisors' assessment of their teaching skills focused primarily on content pedagogy. Even in general, university supervisors were reported to offer very little feedback in this area, and what feedback they did offer was often in terms of how to "simplify complicated subjects."

Master teachers, on the other hand, were perceived to offer "a lot" of feedback in this area (although less than 25% of the teachers said it was the primary focus of their master teachers' assessment of their skills). Elementary teachers talked of receiving feedback from their master teachers about how to teach science, math, and literature in particular. Secondary teachers recalled that their master teachers taught them "specific techniques for teaching specific concepts" or "critiqued their lessons," offering suggestions "on how to make them more interesting." One middle school teacher commented:

> *I received a lot of feedback from my master teacher about how to teach English. I pattern my class now directly after my master teacher's class.*

During hiring interviews, the questions the teachers reported being asked about their skills in teaching specific subjects varied tremendously. On the one hand, some teachers were asked very specific questions related to their subject matter such as "How would you describe your lab routine?" or "If the children were studying the Gold Rush, how would you present it?" One high school English teacher, who recalled being interviewed by two English teachers and a history teacher, commented:

> *I was asked a lot of questions about my skills in teaching English. What is the writing process? How would I teach an essay? A lot of questions were on teaching writing. What thesis statement would I put with this book?*

Other teachers, however, recalled being asked more general questions: A teacher being interviewed for a middle school English position was asked, "How would you feel about teaching English?"; a teacher being interviewed for a high school social studies position reported being asked, "What can you teach? What are you comfortable teaching?" Overall, although most of the teachers recalled being asked at least one question about their skills in teaching a particular subject, over a third of the teachers (8 of 22) said their hiring interviews did not include any questions about their skills in teaching a particular subject.

During the first year of teaching, very few teachers reported receiving feedback about their skill in teaching a specific subject. While three of the 22 teachers said they were told that their skill in teaching a specific subject was a strength, not one of the teachers said this area was identified as a weakness, either after the first or last assessment.

In summation, content pedagogy never seemed a primary focus in any of the assessments that the teachers experienced, beginning with student teaching. According to the teachers, university supervisors spent little time assessing this area, while master teachers spent much more time and offered more information and helpful ideas. During hiring interviews, although some teachers recalled being questioned extensively in this area, most did not. During the first year of teaching, content pedagogy seemed even less of an assessment focus as none of the teachers said content pedagogy was identified as an area of weakness for them, and few teachers said it was identified as a strength.

Working with Diverse Students

Although some may say that a teacher's skills in working with diverse students can be subsumed under the category, General Pedagogy, we believe that student diversity is one of a teacher's main challenges in today's classroom and thus warrants a special look at how it is addressed by the current assessment system. As was done with the other content areas, we began by asking the teachers about the feedback they received in the area of working with diverse students (e.g., students of different ethnic groups, different ability levels, limited-English proficient students) from their university supervisors and master teachers during student teaching. Based on the teachers' responses, this area played a very small part in the content of their student teaching assessments. For example, over half of the teachers said their university supervisors gave no feedback at all in this area or "very little." Of those teachers who did recall feedback in this area, only three described feedback that directly related to the instruction of diverse students: Two teachers received feedback about the way they instructed LEP students, and one teacher received feedback about the way he introduced a lesson to a particular group of students (he didn't describe the students). Other teachers described feedback that tended to be more oriented towards classroom management, e.g., suggestions about "working with students with social behavior problems," "how to prepare to handle potentially explosive circumstances," "how to improve the room environment for low-income students," or "certain populations need extrinsic rewards."

Master teachers, like the university supervisors, were also perceived as offering little feedback in the area of working with diverse students. The feedback that was provided,

however, tended to differ from that of the university supervisors in one respect: Although master teachers were also recalled giving feedback in the areas of working with LEP students and students with behavior problems, master teachers were more likely, according to the teachers, to focus their feedback on working with students of different ability levels. Teachers reported being given feedback about "different approaches to working with different ability levels," "grouping for reading by ability levels," and "grouping students" in general.

When being assessed for hiring purposes, the majority of teachers said they were asked questions about working with diverse students. These questions ranged from the very general -- e.g., "What experience have you had?" -- to the very specific. One teacher, for example, relayed several questions she was asked about working with LEP students:

> *I was asked, What would you do to help an LEP student to succeed? What is your attitude toward LEP students? Are you aware of the resources available to help LEP students?*

Although the questions asked during the hiring interviews seemed to focus on a wide range of student groups (e.g., different ability levels, different ethnic groups, mainstreamed students, and LEP students), the questions recalled by the teachers regarding working with different ethnic groups tended to be the least specific. A teacher who was applying for a position in a school that was trying to recruit more minorities, was asked, "Have you worked with diverse student groups?" The most specific question recalled by any of the teachers regarding working with diverse ethnic groups was the question, "What ethnic group would you be most comfortable in teaching?" This teacher said that his interviewer told him the differences between the different ethnic groups and then asked him to make his choice (he was not, however, guaranteed a placement with his preferred group).

While the hiring questions revealed a concern by the districts and schools that their beginning teachers be competent in working with diverse student groups, that concern was not revealed by their first-year assessments. That is, while teachers recalled hiring questions that specifically asked about their skills in working with diverse student groups, only two teachers recalled being told after their first or last assessment that they had a strength or weakness in working with diverse student groups. A Spanish bilingual high school teacher was told he "was good for the kids he was serving," and an inner city high school teacher whose classes were primarily filled with low-ability Hispanic students was told she had a "good knack with these types of kids" and did a good job "getting them concerned about their grades." None of the other teachers, however, said that as a result of

their assessments they were identified as having strengths or weaknesses in the areas of working with LEP students, students of different ability levels, mainstreamed students, etc.

In summation, working with diverse students seemed to play a minor role in the assessments the teachers experienced in student teaching through their first year of teaching. Most teachers perceived their university supervisors as giving very little attention to this area -- unless it dovetailed with classroom management issues. Master teachers were also perceived as giving this area low priority, although a fair number of master teachers were credited with directing their student teachers' attention to working with students of different ability levels. Teachers' recollections of hiring questions revealed that districts and schools are concerned about student diversity, but the depth of that concern varies tremendously from district to district or school to school. Moreover, during their first year of teaching, almost none of the teachers had strengths or weaknesses identified in this area after either their first or last assessment, indicating again that this area is not a major focus of assessment.

Curriculum Frameworks

The content of the state's curriculum frameworks varies, of course, depending on the subject being discussed (e.g., English/language arts, mathematics, science). We wondered, however, if the content of the curriculum frameworks was reflected in any way in the content of the assessments the teachers experienced in student teaching through their first year of teaching. In particular, we interviewed the teachers about any questions they may have been asked in their hiring interviews about their knowledge of the state frameworks. Over half of the teachers could not recall being asked any questions about their knowledge of the state frameworks. In fact, only six teachers recalled being asked a specific question or questions about their knowledge of the frameworks. Some of the teachers' comments are as follows:

> *I was asked a couple of questions concerning the [science] state curriculum. These were passing questions, not an area of concentration.*
>
> *He asked me about the trends in the English framework.*
>
> *They asked me some questions that led to a literature-based approach discussion.*
>
> *She wanted to know if I was familiar with the new social science framework?*

One of the teachers who was not asked any questions, offered a possible explanation:

> *The state curriculum presupposes that kids have been in the pipeline for years and these students haven't. Perhaps that's why he didn't ask, because the framework needs modification for the Mexican immigrant population in class.*

The teacher's explanation is a possible one; other possibilities, however, suggest that the content of the curriculum frameworks is not a major focus of assessment at the point of hiring or that subject matter content is not always identified with the frameworks (e.g., a question might have been asked about working with math manipulatives but without linking it to the content of the math framework). Whatever the explanation, the lack of assessment in this area corresponds with the lack of hiring questions that were asked about content pedagogy as discussed above.

As for the other assessment points, none of the questions we asked referred directly to the curriculum frameworks. Looking at the teachers' descriptions of the feedback they received from their university supervisors' and master teachers' assessments, however, we can note that not one of the teachers directly referred to the curriculum frameworks in their responses. A few teachers made indirect references by responding, for example that their master teachers gave them feedback about math manipulative use or hands-on science activities, two techniques mentioned in the math and science frameworks respectively. Similarly, reflecting on the assessments made during their first year of teaching, none of the teachers made any specific reference to the curriculum frameworks when describing their strengths and weaknesses as they were identified by the assessments, nor were any indirect references made which could be linked to the curriculum content.

In summation, the state curriculum frameworks and their respective contents seem to play an extremely minor role in the assessment of beginning teachers. Although some teachers recalled being asked questions during their hiring interviews which assessed their knowledge of the state curriculum frameworks, most did not. Moreover, when describing information they received at other assessment points--i.e., during student teaching and in the first year of teaching--none of the teachers ever made a direct reference to the curriculum frameworks and only a few made an indirect reference to the framework content. (This may be a reflection of that fact that the primary focus for most beginning teachers in their first few years is survival; curriculum framework concerns may come later when teachers feel more secure about their general pedagogical skills.)

Summary and Conclusions

The following summary and conclusions are based on the beginning teachers' interview responses.

The Forms of Assessment

1. In general, assessment of a beginning teacher's knowledge, skills and abilities primarily takes two forms: Written exams and observations.

   - Written exams are used to assess a prospective teacher's knowledge (e.g., subject matter knowledge) and basic skills before the start of student teaching.

   - Observations are relied on to assess a beginning teacher's knowledge, skills and abilities during student teaching and in the first year of teaching.

Characteristics of Assessors

1. Based on the teachers' responses, the majority of assessors could be perceived as either relatively inexperienced or in some way inadequate (e.g., lacked the same subject matter background) as assessors.

   - The majority of master teachers were perceived as having had five years or less supervisory experience.

   - One third of the master teachers were thought to be supervising a student teacher for the very first time.

   - Almost half of the university supervisors were perceived as having a different subject matter or grade level (e.g., elementary or secondary) background than the student teacher.

Content of Assessments

1. **General pedagogy was the major focus** of all the assessments that the teachers experienced from student teaching through the first year of teaching.

- The pedagogical area most often assessed and remarked upon by assessors was classroom management.

2. **Content pedagogy was never a primary focus** of any of the assessments that the teachers experienced.

   - Master teachers were perceived as providing more content pedagogical feedback than university supervisors.

   - Content pedagogy was assessed more in student teaching than in the first year of teaching.

3. Working with diverse students seemed to play a minor role in all the assessments the teachers experienced.

   - University supervisors were perceived as giving little attention to this area unless it dovetailed with classroom management.

   - Master teachers were most likely to focus assessment in this area on working with students of different ability levels.

   - The district and school hiring questions recalled by the teachers reveal a concern about student diversity, but little depth.

   - With the exception of establishing rapport with students, working with diverse students was an area rarely identified as a strength or weakness in any of the assessments experienced during the first year of teaching.

4. The state curriculum frameworks seem to play an extremely minor role in the assessment of beginning teachers.

   - Except when they were describing how they were hired (e.g., the questions that were asked), no direct reference and few indirect references were ever made to the curriculum frameworks during the teacher interviews.

Rigor of Assessments

1. Based on teacher responses, rigor of an assessment depended on the amount of areas being assessed, the amount of work expected, and the level of expectations held by the assessor(s).

2. From the perspective of the beginning teachers in the study, the assessments experienced from the point of admission to a teacher education program through the first year of teaching **are not rigorous** and demanding.

   - Less than 50% of the teachers described any of the assessments they experienced as rigorous and demanding.

   - Only two teachers described as rigorous all the assessments they experienced from admission to the teacher education program through the first year of teaching (see Table 5.4).

   - Almost none of the teachers found the assessments conducted for admission to student teaching to be difficult.

3. Overall, student teaching assessments were considered more rigorous than assessments experienced during the first year of teaching.

Thoroughness of Assessments

1. Based on teacher responses, **the assessments were not thorough** or comprehensive.

   - During student teaching and in the first year of teaching, **teachers wanted more observations, more feedback, and a greater variety of assessors**. Some teachers also wanted observations that lasted longer than the typical time of one hour.

   - In the first year of teaching, 25% of the teachers said they had no weakness identified after their first evaluation, and a majority of the teachers had no weakness identified after the final evaluation.

Redundancy of Assessments

1. There is **much redundancy in the focus** of the assessments from student teaching through the first year of teaching.

   - From student teaching through the first year of teaching, assessors are perceived as primarily focusing on a teacher's general pedagogical skills, especially classroom management.

   - A little over 40% of the teachers reported having a same strength identified after their first student teaching assignment as after their first evaluation in their first year of teaching, suggesting that there may be some duplication of assessment.

Consistency of Assessments

1. During student teaching and in the first year of teaching,the assessments were **consistent in their focus** on general pedagogy.

   - Based on teacher responses, assessors consistently looked for evidence of good classroom management, good student rapport, and to some extent, good ways of involving students in activities.

2. There was a **definite lack of consistency in the frequency and timing** of assessments experienced during student teaching and in the first year of teaching.

   - The range in frequency of university supervisor assessments was two observations in 18 weeks to one observation every week.

   - Secondary teachers reported being observed by their university supervisors in some classes but not in others.

   - Observations by all assessors ranged from 15 minutes to one hour.

   - Some teachers reported daily feedback from their master teachers; others received almost no feedback at all.

- The timing of the first assessment made in the first year of teaching ranged from the first month of the school year to sometime after January.

3. Student teaching assessments were not perceived as consistent in the identification of strengths and weaknesses.

   - Almost all of the teachers said that the evaluation of their strengths and weaknesses made at the end of their first student teaching assignment was different than that made at the end of their second student teaching assignment.

   - Teachers reported having different strengths and/or weaknesses identified by the assessments. Some teachers reported that a strength in one evaluation became a weakness in the next.

Fairness of Assessments

1. Based on teacher response, **most assessments were fair**. (Fairness did not seem to be an issue for minority teachers: Of the eight minority teachers in the sample, only one described an assessment as being unfair, and this was because the "end evaluation did not correspond to daily feedback.")

Helpfulness of Assessments

1. Based on teacher response, **most assessments were helpful in some way.**

   - During student teaching, the assessments were most commonly perceived as helpful because of feedback provided in the area of general pedagogy. University supervisors' assessments were usually valued for the help they provided in classroom management.

   - First-year assessments were generally perceived as helpful because they provided a feeling of support and encouragement and an increase in confidence rather than specific ways to improve teaching.

2. Overall, the teachers perceived their master teachers' assessments were to be more helpful than the university supervisors'.

- Fewer teachers pointed out deficiencies in their master teachers' assessments than they did in their university supervisors' assessments.

- In general, master teachers were valued over the university supervisors because they saw the student teachers more frequently, were perceived as being more familiar with the classroom and the students, gave more practical, specific, and frequent suggestions, and modelled specific techniques and behaviors.

- In addition to providing help in the area of general pedagogy, master teachers' assessments were also valued by one third of the teachers for the subject-matter help received.

- University supervisors' assessments were perceived as being especially lacking in the areas of content pedagogy (i.e., subject matter help) and working with diverse students.

3. Student teaching assessments were perceived as more helpful than assessments experienced in the first year of teaching.

   - Student teaching assessments were valued over first-year assessments primarily for one reason: more specific and frequent feedback about one's teaching skills is given during student teaching.

4. While most teachers found the assessments they experienced during student teaching and in their first year of teaching to be helpful in some way, many did not.

   - Approximately 25% of the teachers saw no value in their university supervisors' assessment of their teaching skills; approximately 20% saw no value in their master teachers'.

   - Only slightly more teachers found the final evaluation made during their first year of teaching to be helpful than did not.

Assessments and Support

1. The assessments in student teaching and the first year of teaching were perceived by many teachers to be a source of support.

- Many teachers reported that feedback from their master teachers and university supervisors not only provided them with instructional ideas, but also helped boost their confidence in themselves as teachers.

- The majority of teachers who considered their first evaluation in their first year of teaching to be helpful did so primarily because it gave them confidence and a feeling of being supported.

2. Based on teacher responses, the level of support that was provided in connection with the first-year assessments varied greatly.

   - Most frequently, the same person who conducted the assessment, also provided suggestions as to how to address any identified weaknesses.

   - Approximately 60% of the teachers said they received some guidance and assistance from a mentor teacher or other experienced person. (More than half of these teachers were participants in a California New Teacher Project.)

   - After the first assessment, two teachers with identified weaknesses reported they were told to observe other teachers, but only one of the teachers was provided coverage to do so.

CHAPTER 6:

SUMMARY OF FINDINGS ACROSS SUBSTUDIES AND CONCLUSIONS

CHAPTER 6:

SUMMARY OF FINDINGS ACROSS SUBSTUDIES AND CONCLUSIONS

Information from assessments of prospective and beginning teachers can fulfill multiple purposes. Useful feedback on teaching skills allows student and beginning teachers to take pride in their strengths and to pinpoint weaknesses which need special attention for further development. Feedback tailored to a teacher's developmental stage could communicate that one can do well as a student or beginning teacher and still need further development. Assessment information can assist support providers by identifying the most important areas where help and improvement is needed. Assessment information is used to make critical decisions in the areas of admission to a teacher credentialing program, recommendation for a credential, hiring, and retention. If assessment information were systematically collected about graduates and provided to credential programs, it could provide guidance for program improvement. Assessment information also has the potential of assisting the State in designing programs and prioritizing expenditures.

This final chapter draws from the material presented in the preceding chapters to present conclusions regarding the extent to which present assessment practices provide the information necessary to fulfill the purposes described above. It begins by describing the content of assessments, or what is -- and is not -- currently assessed. It next describes the forms of assessment used, or how prospective and beginning teachers are being assessed. It goes on to summarize findings about the technical quality of present assessments, or how well teachers are assessed. It concludes by comparing the collection of assessment practices found in our sample to the characteristics of an ideal assessment system.

Focus of Assessments

The following discussion addresses how major domains of teaching knowledge, skills, and abilities are addressed at the various assessment points.

Subject-matter Knowledge

Subject-matter knowledge, which provides a foundation for the development of skills in content pedagogy, is the primary focus of the assessments conducted at the point of entry into the teacher education program. Assessment is generally conducted in one of two ways:

a standardized examination, approved by the State, or examinations in the courses prescribed by a subject-matter waiver program. Subject-matter knowledge seldom gets reexamined at later points of assessment, as IHEs and LEAs rely heavily on this initial screen. (Most districts, for example, assess subject-matter knowledge by looking at course transcripts.) Subject-matter knowledge may be indirectly evaluated when IHEs and LEAs look at content pedagogy.

The current methods of assessing subject-matter knowledge generally do not identify deficiencies of the type observed in the teachers who participated in many of the CNTP assessment pilot tests. These deficiencies include inadequate knowledge of the structure of the discipline, an inability to sequence topics in a meaningful way, and lack of ability to apply basic discipline constructs (e.g., fractions in math, the different stages of the writing process.) The new subject-matter examinations currently being developed by the State include elements of application which may address these last deficiencies.

General Pedagogy

General pedagogy, especially classroom management, is the major focus of all assessments that the teachers experience from student teaching through the first two years of teaching. The State relies on IHE assessments of general pedagogy for credentialing requirements.

Content Pedagogy

Looking across the assessments conducted by the IHEs, LEAs, and the State, evaluations of a teacher's skills in content pedagogy are consistently lacking in depth. During student teaching, content pedagogy is minimally assessed through observations and examination of lesson plans, and standards tend to be relatively low. The low standards do not seem to be grounded in a clearly defined perspective on teacher development. Many aspects of content pedagogy such as anticipating student errors and choosing forms of instruction that are best suited to specific students largely depend on experience-based knowledge, suggesting that standards for student teachers should focus on evidence of progress toward developing these competencies in the topics taught. As with general pedagogy, the State again relies on IHEs to evaluate content pedagogy for credentialing purposes.

Content pedagogy is also minimally assessed by LEAs at the point of hiring and/or in the first year of teaching. In secondary schools, for example, the department chair may

participate in the hiring interview(s) and/or assist in the evaluation for continued employment. In schools where teaming is emphasized, either in actual instruction or in planning, a teacher's skills in content pedagogy may also be evaluated informally by peers. Skill in content pedagogy is most often defined by the LEAs as the ability to use the instructional techniques emphasized by the district.

Knowledge of Students and Student Diversity

Knowledge of students and of teaching diverse students is generally a very minor focus of all the assessments that a teacher experiences from the point of student teaching through the first year of teaching. Only a small amount of feedback during student teaching is focussed on teaching diverse students, and tends to be directed toward classroom management issues or teaching students of differing ability levels. Knowledge of students is most directly assessed by LEAs at the point of hiring, when specific questions are sometimes asked about a teacher's familiarity or experience with certain types of students.

General Academic Ability and Personal Attributes

These are always assessed. General academic skills are evaluated at every entry point, usually through multiple measures: a mandatory standardized examination (CBEST), GPA, and occasionally through performance assessments (e.g., a writing sample). Personal attributes are primarily assessed through observation and interviews. Although IHEs and LEAs differ in some of the specific attributes they emphasize, almost all look for enthusiasm, rapport with children, and a commitment to teaching.

Forms of Assessments

The skill areas described above are assessed by IHEs, LEAs, and the State using various forms of assessment. These forms and their uses include:

- **Document review** is one of the most common forms of teacher assessment used at points of entry. IHEs and LEAs vary, however, in how thoroughly they review materials and in the types of inferences they make from the same information.

- **Written examinations** are used by IHEs and the State to assess subject-matter knowledge and basic skills proficiencies. IHEs also use written examinations to assess general and content pedagogy, usually before the start of student teaching.

- **Observation** is the primary form of assessment used by IHEs and LEAs to assess a beginning teacher's knowledge, skills, and abilities during student teaching and in the first two years of teaching.

- **Alternative performance assessments** (e.g., developing unit and/or lesson plans, portfolios) are secondary forms of assessments sometimes used during teacher education courses, student teaching, and the first two years of teaching. The State is moving to incorporate performance assessment into the subject-matter knowledge examinations.

- **Interviews** are commonly conducted by IHEs at the point of entry into a teacher education program and by LEAs at the point of hiring.

- **Alternative methods of assessment** such as measuring student achievement and considering student feedback (i.e., the students' opinions of the teacher's performance) are occasionally used during student teaching and in the first year of teaching.

Technical Quality of Assessments

The technical quality of these assessment practices varies considerably both between and within institutions, but, with the exception of the standardized State assessments, is generally low.

Clearly stated expectations and performance standards are rarely a characteristic of present assessments. Moreover, the assessments with the clearest criteria and standards are those that measure general academic ability and basic skills proficiencies; they do not measure teaching skills. Criteria used in present assessments addressing teaching skills are most often represented by broad categories whose interpretation is largely left up to individual assessors. Rating scales, when used, are rarely defined. In some small credential programs and districts, frequent conversations among the limited number of assessors increases the uniformity in interpretation and rating of criteria. In larger programs and districts, extensive training would be needed to achieve the same result.

Not surprisingly, the lack of clear expectations and standards results in preparation of assessors that, when it occurs, is largely oriented more to procedure than to substance. Even this extent of preparation varies by assessment point. For instance, training provided by districts with respect to assessments for retention is common, while training for

assessors conducting hiring assessments is not. Although the sample of beginning teachers indicates that most of the detailed and frequent feedback received during student teaching is from cooperating (or master) teachers, the cooperating teachers themselves tend to receive no training for this role, and over one-third of the cooperating teachers interviewed reported some confusion over the meaning of some parts of the form which they used to rate student teachers.

The sample of beginning teachers also indicated a wide variance in the frequency and timing of the assessments of their teaching skills that could not be explained by the number of problems they may have been experiencing. Some teachers were observed frequently; others were not. Some teachers were observed beginning early in their teaching assignment; others were not. This was true of assessments both during student teaching and during the beginning years of teaching.

Extent to which IHE, LEA, and State Assessments Constitute an Assessment System

The guiding constructs for an ideal statewide system, described in the introduction, are rigor, thoroughness, fairness, consistency, and helpfulness. The extent to which these constructs characterize the assessment practices identified in the overall study is discussed separately for each construct.

Rigor

Both IHE and LEA assessment practices are uneven in terms of rigor, both across institutions and within institutions and credential programs. Evaluation criteria are identified, and sometimes defined by detailed examples. However, standards, i.e., the rating categories used, are rarely defined.

Assessment standards applied are frequently quite low, and few documents examined reflected a well-articulated perspective on the development of teaching skills. The evaluation form and procedures for evaluation of teachers at an early stage (e.g., student teacher in first placement, first-year teacher) was almost always the same as that for teachers at later stages (e.g., student teacher in second placement, twenty-year veteran). Assessors reported that they took the extent of a teacher's experience into account in their evaluation, but the basis for doing so was not clear, and in many cases, appeared to be at the discretion of the individual assessor. Only one form for evaluating student teaching reflected a developmental perspective by indicating that some teaching skills would not be

observable at first, and the number of skills rated would increase as the student teacher gained more experience.

LEAs differ in their capacity to maintain high standards in the evaluation of prospective and beginning teachers. Standards for assessment of beginning teachers by LEAs are driven by market forces, where beginning teachers are judged relative to those who might be expected to replace them. Districts vary in their attractiveness to teachers, and some of the most influential reasons such as an inner city location or pupil funding levels are beyond the district's control. Although recruitment strategies have some impact, a district's attractiveness strongly influences the number and quality of applicants who apply. As a result, highly selective districts are able to apply high standards, while districts with difficulty recruiting credentialed teachers are not. However, districts with severe staffing problems *do* employ standards by which they reject credentialed teachers. These standards often include more emphasis on knowledge of and empathy with the type of students in the district and less emphasis on pedagogy, reflecting a belief that pedagogy is more easily taught.

For assessment methods which rely on the professional judgment of the assessor (e.g., observation, interviews), the amount of training of assessors varies considerably. Outside of State assessments, most training focusses more on the *process* of assessing rather than on the application of specific assessment criteria or standards. Those IHEs and LEAs that define the constructs being assessed, either conceptually or by providing examples, usually have the most rigorous assessments as well.

In terms of identifying candidates who should be eliminated from the profession, the assessment is most rigorous at the entry points. From the perspective of the beginning teachers interviewed, the rigor of assessment depends on the variety of areas being assessed, the amount of work expected, and the level of expectations held by the assessor(s). Most considered their student teaching assessments to be more rigorous than those experienced during the first year of teaching.

The rigor of assessments was affected by resources and the methodology available. State assessments exhibit high technical quality, in part because the high cost of well-developed assessments can be spread across a large number of teacher candidates, and a source of income is available to fund administration, development, and implementation costs.

For eliminating candidates from the system, IHEs and LEAs tend to favor assessments that are "appeal-proof," which means one of two characteristics. The first characteristic is that the assessment criteria are directly measurable by observation. For example, while the validity of using a GPA as a measure of academic skills may be argued, whether or not a specific GPA meets the standard cannot. However, whether or not a teacher is using an appropriate initiation to a lesson or is using an appropriate discipline technique is less clear, and more vulnerable to conflicting interpretations. The degree to which indicators of skills can be clearly observed varies across teaching domains. Classroom management is relatively easy to observe and document, while content pedagogy and knowledge of students are not. LEAs are not required to give reasons for releasing teachers prior to receipt of tenure, and assessors in many of the districts in the sample stated that they did not explain decisions to not rehire beginning teachers to avoid potential legal problems. Some districts even have recommended guidelines based on legal criteria as to what assessors should and should not tell teachers at the point of dismissal.

The second characteristic influencing the ability of a negative evaluation to withstand appeals is the degree to which there is a consensus on a definition of important teaching skills across contexts and teaching styles. One reason that content pedagogy and knowledge of students are difficult to document is that they are difficult to define across different teaching contexts and styles.

Thoroughness

Assessments seem to be conducted without the benefit of a general teaching framework which serves as a context in which to discuss a beginning teacher's knowledge, skills, and abilities. Often there are lists of competencies, but it is not clear how the knowledge, skills, and abilities identified interrelate. The lists are also often defined at a level that either stops at the description of the teaching domain (e.g., "assessment of student progress") and/or which seems to be more characteristic of experienced than student or beginning teachers (e.g., "paces individual lessons appropriately" for a student teacher). Without a general teaching framework, each teacher is left to generalize from their particular teaching experiences and to invent their own teaching framework. As a result, the learning curve for many teachers is considerably lengthened. Prospective and beginning teachers seem to have difficulty in putting feedback about specific actions and reactions into a larger framework so that they see alternative ways of approaching problems. The lack of a framework also facilitates the rejection of negative feedback about specifics because the teacher is unable to link the feedback to a larger principle or issue. Thus the feedback is

more vulnerable to being ignored, or seen as a result of a personality conflict, or a lack of understanding by the assessor of a particular context.

Information from other assessment points tend to be underutilized, even within the same organization. LEA use of information from IHEs is limited. Many districts collect records of student teaching assessments. However, where the district controls the initial selection process, the information is not routinely provided to school-level personnel who make the final hiring selection. Even where information is available, a lack of time to consider the information means that it most often either is not used or is used only in the final stages of the selection process, where the number of active candidates is considerably reduced.

Information from previous assessment points is rarely used for diagnostic purposes, both across institutions (i.e., from the IHE to the LEA) and within institutions (e.g., information from IHE education coursework and student teaching; formative and summative evaluations of teachers in LEAs). Even if assessors or institutions are willing to share and use information, there are legal liabilities with respect to negative information which adversely affect the willingness to share information. Generally, only "fatal" weaknesses of applicants are disclosed. Minor to major weaknesses are often identified by omission, requiring an evaluator to go well beyond what is written in letters of recommendation. Typical comments were: "Unless the letter of recommendation specifically states that classroom management is strong, we assume that it isn't," and "Listen to what they *don't* say as much as to what they *do* say."

Redundancy in assessments does not seem to be a problem. Skills vary in the extent to which they are reassessed at different assessment points. Subject-matter knowledge is assessed both in relevant coursework and either at the point of entry to a teacher credential program, for graduates of California credential programs, or at the point of application for a credential, for other graduates. Generally, IHEs and LEAs trust these initial assessments. General pedagogy is assessed at all points from student teaching on, as is appropriate, since expertise in this skill area can be expected to increase over time or be applied in a different context. However, the documents and forms examined do not indicate increased expectations over time within the same context, relying on the individual assessors to alter expectations accordingly.

Fairness

The absence of a conceptual framework of teaching knowledge, skills, and abilities makes it difficult to set fair levels of expectations. Because of the lack of explication of assessment standards, fairness often depends on the professional judgment of people who are not always well trained. The differences in teaching contexts, especially for IHEs, also contribute to the difficulty of maintaining fairness in application of standards, perhaps contributing to the reluctance to set standards.

Some attention to underrepresented groups occurs at the entry points. IHEs can admit up to 15% of their students who don't meet the GPA requirement, and some LEAs reported giving a preference to underrepresented groups at the point of hiring. Once these teachers are admitted or employed, however, they must meet the same standards as other teachers. The greatest barrier to increasing the number of underrepresented groups in teaching is the basic skills proficiencies requirement.

Consistency

There are some criteria which are commonly applied across organizations at the same assessment point. IHEs are guided by the CTC program quality standards which specify skills to be demonstrated by program graduates. LEAs must assess teachers by criteria which include those specified in the Stull Act. However, interpretation of the criteria and standards applied vary both within and across institutions. Some other common criteria also appear across assessments. Liking children is commonly employed in IHEs, while enthusiasm is a common criterion in LEAs. The absence of a common conceptual framework of teaching skills inhibits the application of consistent criteria, as does the minimal training often provided for assessors. Some small credential programs and districts achieve consistency through the frequent interaction of assessors.

The teacher descriptions of their assessments revealed a definite lack of consistency in the frequency and timing of assessments, both during student teaching and in the first year of teaching. The range in frequency of assessments by university supervisors was two observations in eighteen weeks to one per week. Some teachers reported daily feedback from their master teachers, while others reported receiving almost no feedback at all. The timing of the first assessment made in the first year of teaching ranged from the first month of the school year to some time after January. The duration of observation by assessors ranged from 15 minutes to one hour.

Helpfulness

Helpfulness is directly related to the extent of the formative feedback provided by an assessment. Except for the final evaluation, student teaching evaluations are supposed to be formative in nature. The beginning teachers in the sample most commonly reported these assessments to be helpful; however, approximately 25% of the teachers saw no value in their university supervisor's assessment, and about 20% felt the same way about their master teacher's assessment. Master teachers were generally valued over university supervisors for providing feedback because they saw the student teachers more frequently and were perceived as being more familiar with the classroom and students. The sample teachers also reported that master teachers gave more practical, specific, and frequent suggestions, compared to university supervisors, and modelled specific techniques and behaviors. General pedagogy was the area in which most teachers reported receiving useful feedback. First-year assessments were generally perceived by the teachers to be helpful because they provided a feeling of support and encouragement and an increase in confidence, rather than because they provided specific suggestions to improve teaching. Only a slight majority of teachers reported the final evaluation made during their first year of teaching to be helpful.

Assessors in most districts described attempts to avoid overwhelming the new teacher by identifying problems. Weaknesses identified at hiring were often communicated as suggestions for staff development activities. Communication of problems seems difficult for assessors. Many spoke of the need to preserve the self esteem of the beginning teacher, who faces a considerable workload and stress. However, the result is that the negative feedback that teachers currently get, both in student teaching and in the first year of teaching, is often couched in the form of helpful suggestions (e.g., "Have you thought of trying..."), which do not necessarily identify weaknesses. Some teachers are unaware of their weaknesses because they are not specifically labeled as such.

In the context of a shortage of teachers and a profession where it is gradually being acknowledged that teaching skills can take years to develop, the best assessment system does little good to identify mediocre teachers unless the teachers are supported in improving their skills. This support should include not only help in improving deficiencies, but also support in maintaining and/or developing strengths. Every district in our sample provided some form of support for beginning teachers. At least seven districts reported assigning beginning teachers to a specific support provider, and 65% of the teachers reported receiving guidance and assistance from an experienced teacher. However, the support providers in six of the seven districts clearly attempted to separate the evaluative role from support by observing only reluctantly in classrooms or observing only when requested to do so. Most

support providers interviewed reported frustration at having little time to support beginning teachers. Many are not released from any regular teaching duties. Others were provided with minimal release time (e.g., one extra preparation period for high school teachers), but were assigned to multiple teachers.

Although support providers tend to express concerns about the ability of beginning teachers to hear negative feedback, the beginning teachers interviewed expressed a desire for more feedback. They prefer that it be specific (e.g., "Your initiation to the lesson captured the attention of almost all your students") rather than general (e.g., "You're doing fine"). They would also prefer to receive feedback from more than one person.

Conclusions

The assessment practices documented by the overall study do not closely resemble an ideal assessment system. It is questionable whether the assessment practices conducted by the State, seventy-three institutions of higher education, and over 1,000 districts will ever perfectly match the ideal. However, the uncoordinated patchwork of existing legislation which mandates assessment practices and the lack of a generally accepted framework to describe effective teaching probably widen the disparity in assessment practices across institutions.

Rigorous evaluation of many aspects of teaching takes time and qualified assessors. Both seem to be in short supply for many institutions. Furthermore, the number of teachers to be assessed by IHEs and LEAs is large. When extremely large numbers of teachers must be assessed, the time available per teacher for assessment is significantly reduced, since supervision is labor intensive, and institutions seldom can increase their capacity proportionately. It also affects assessment standards and the degree of assistance provided. A sort of "triage" system is practiced where the most needy get the most help, and those doing okay are left alone to cope as best as they can. Although all new teachers need assistance, the system relies on teachers performing adequately to develop on their own, and is highly unlikely to push these teachers to further develop.

The findings of the overall study do suggest actions which could improve existing assessment practices and move the totality more toward an ideal system:

- **Creation and adoption of a state teaching framework.** A state teaching framework should be created that emphasizes the developmental nature of teaching. The framework should represent a broad consensus and clearly

identify the knowledge, skills, and abilities expected of beginning teachers and those teachers with more experience. The framework should be broadly defined so that it applies to different teaching contexts and subject areas. Such a framework should recognize that IHEs and districts will vary in the degrees to which different skills are emphasized and that exhibition of the skills may look different in different contexts and subjects. Once the framework is developed and adopted, it may suggest changes in existing legislation to more tightly coordinate evaluation efforts.

- **Increase in assessor training.** Training for assessors needs to be increased, particularly in the area of methodologies to evaluate content pedagogical skills. Although individual assessor differences would continue to occur (and would be accompanied by context differences), training in evaluation of knowledge, skills, and abilities contained in a teaching framework might increase the variety of teaching skills examined and increase consistency of evaluation across teachers. In particular, both IHE and LEA assessors appear reluctant to evaluate skills rigorously in content pedagogy because of an inability to document deficiencies in a clear, consistent way. The increased training would need to be accompanied by increased time for evaluation, as all of the assessors interviewed seemed to be working to capacity now.

- **Increase in the scope of teaching assessments.** At the present time, many teachers do not receive feedback on the full range of teaching skills. Some teaching skills, particularly those in the area of content pedagogy and knowledge of students and student diversity, are not well assessed. Prospective and beginning teachers should receive feedback on a broad range of teaching skills and receive assistance for improvement of these skills.

- **Increase in formative feedback provided to teachers.** Teachers were eager for more feedback on their teaching, particularly if it is accompanied by suggestions for improvement, assistance for doing so, and some means of protection from personality conflicts and assessor errors. Ideally, such feedback should be provided by more than one person and be frequent enough to be helpful.

These proposals would have the effect of not only increasing the capacity of local school districts to conduct evaluations, but also connect these reforms with other reform efforts such as restructuring and the implementation of the new curriculum frameworks. At a minimum, the adoption of these changes would increase the likelihood that each student

teacher and beginning teacher would hear a common language spoken about their expected skills and knowledge. Their journey from student to novice teacher would be made smoother by supervision from articulate, trained assessors who espouse a common vision of teaching. In turn, the widespread implementation of these reforms would enhance the education of future generations of California children.

BIBLIOGRAPHY

BIBLIOGRAPHY

Commission on Teacher Credentialing (1990). **Standards of program quality and effectiveness, factors to consider and preconditions in the evaluation of professional teacher preparation programs for multiple and single subject credentials**. Sacramento, CA: Commission on Teacher Credentialing.

Educational Testing Service (1991). **CBEST. California Basic Skills Educational Skills Test.** 1991-92 Information Bulletin. Oakland CA: Educational Testing Service.

Majetic, Richard (1990). **The California Basic Educational Skills Test: Annual report of examination results**. Sacramento CA: Commission on Teacher Credentialing.

Majetic, Richard (1990). **California Basic Educational Skills Test: Analysis of 1985 examinees and those who repeated the examination in subsequent years.** Sacramento CA: Commission on Teacher Credentialing.

Watkins, Richard (1985a). **Practitioner review of the California Basic Educational Skills Test specifications, content, and difficulty**. Sacramento, CA: Commission on Teacher Credentialing.

Watkins, Richard (1985b). **Third year passing rates on the California Basic Educational Skills Test (CBEST) and passing rates by institution attended**. Sacramento, CA: Commission on Teacher Credentialing.

Wheeler, Patricia (1988). **A study on the appropriateness of fifteen NTE specialty area tests for use in credentialing in the State of California. Volumes I and II.** A report prepared for the California Commission on Teacher Credentialing by Educational Testing Service, Emeryville, CA.

Wheeler, Patricia and Patricia Elias (1983). **California Basic Educational Skills Test: Field test and validity study**. Berkeley, CA: Educational Testing Service.

APPENDIX A:

SAMPLE CBEST ITEMS

BEST COPY AVAILABLE

Sample Reading Items

Reading Passage:

In the 1850's most of the thousands of Chinese men who bought passage on cramped and often unhealthy ships and journeyed to the "Golden Mountain," as they referred to California, obtained work in the gold-mining fields. There they exhibited not only formidable endurance but also skill and ingenuity.

The pioneer Chinese miners excelled at handling the rock pan, a device used to separate gold from lighter materials as river water washed through it. They diligently mined claims that others believed were exhausted, recovering great quantities of gold that would otherwise have been abandoned. Drawing on their experience with agricultural irrigation in their homeland, the Chinese also contributed technologies previously unknown and unavailable in the mining fields. For example, they worked in teams to construct pine log wing dams several hundred yards long to divert the river flow from gravel bars so that miners wielding picks could get at the exposed soil. And to the astonishment of the non-Chinese miners who had never seen anything of the kind, they introduced treadle pumps that swiftly cleared the water out of mine holes.

1. (**Logical comprehension**): The central idea of the passage is that the pioneer Chinese miners in the 1850s:

   (A) were the only group of immigrants to work in the California goldfields.
   (B) endured hardships on their voyages from China to the "Golden Mountain."
   (C) earned a good deal of personal wealth from gold mining.
   (D) contributed distinctive talents and engineering expertise to mining in the California goldfields.
   (E) differed from other miners in their use only of traditional Asian mining methods to extract gold.

2. (**Inferential comprehension**): It can be inferred from the passage that which of the following was a major problem faced by miners in the early California goldfields?

   (A) The gold had to be extracted from sites covered by water.
   (B) Antagonism existed between groups of miners using different mining techniques.
   (C) Many gold-rich mining sites suddenly became exhausted.
   (D) Gold had to be moved a great distance from the remote places where it was mined.
   (E) Miners in the field had inadequate supplies of food and tools.

3. (**Critical comprehension**): The main purpose of this passage is to:

   (A) review the cultural attitudes of the pioneer Chinese miners.
   (B) explain the high degree of organization among the pioneer Chinese miners.
   (C) underscore the technological resourcefulness of the pioneer Chinese miners.

(D) indicate the abilities of the pioneer Chinese miners.
(E) describe the metallurgical knowledge of the pioneer Chinese miners.

Sample Math Items

1. (**Problem solving processes, arithmetic**): If Donald purchased 98 plants at $0.62 each, what was the total cost of the plants?
   Which of the following methods is most appropriate for estimating the total cost in the problem above?

   (A) 20 X 3 (B) 90 / 0.6 (C) 90 X 0.65 (D) 100 / 0.6 (E) 100 X 0.6

2. (**Problem solving processes, algebraic**): Carol's average bowling score for 3 games was 138, and her highest score for the three games was 24 points higher than her average score.
   Which of the following CANNOT be determined from the information given above?

   (A) Carol's highest score
   (B) Carol's lowest score
   (C) The sum of Carol's two lowest scores
   (D) The sum of Carol's scores for the 3 games
   (E) The difference between her highest score and her average score for the 3 games.

3. (**Problem solving processes, geometric**): A dozen soup cans, all the same size, are arranged in a carton as shown above [shows sketch of three by four arrangement of cans). If each soup can has a diameter of 6 centimeters and is 10 centimeters tall, what are the smallest outside dimensions, in centimeters, of the carton?

   (A) 3 by 4 by 6
   (B) 6 by 6 by 10
   (C) 18 by 24 by 10
   (D) 30 by 40 by 6
   (E) 36 by 48 by 10

4. (**Solving applied problems, algebraic**): Daniel spent $9 on tomatoes and avocados to make a salad that requires twice as many tomatoes as avocados. If tomatoes were 10 cents each and avocados were 25 cents each, how many avocados did he purchase?

   (A) 40 (B) 30 (C) 25 (D) 20 (E) 10

5. (**Solving applied problems, geometric**): If one face of the cube shown above [refers to sketch of cube] has an area of 100, then the volume of the cube is

   (A) 100 (B) 400 (C) 600 (D) 1,000 (E) 2,500

6. (**Mathematical concepts and relation, geometric**): Which of the following computations could be used to find the area of a rectangle that is 8 meters by 12 meters?

   (A) 20^2 (B) 12 + 8 (C) 12 X 8 (D) 2(12 X 8) (E) (2 X 12) + (2 X 8)

7. (**Mathematical concepts and relationships, arithmetic**): What is the least common multiple of 2, 3, 6, and 9?

   (A) 1 (B) 9 (C) 18 (D) 156 (E) 324

Sample Essay Topics

1. "Americans like to think of themselves as individualists but in fact behave most of the time as conformists."

   On the basis of your own experience and observations, explain why you agree or disagree with this quotation.

2. Describe one experience that you had in college that made you a different person from the one you were when you entered.

APPENDIX B:

COMPARISONS WITH PHASE II ASSESSMENTS

Experiences of Beginning Teachers: Comparisons with CNTP Innovative Assessments

In this section of the report, we offer some comparisons of assessment results for six of the beginning teachers in our sample. As stated at the beginning of this report, all of the teachers in the study had participated in at least one of the CNTP pilot tests of innovative assessments that were conducted between 1989 and 1991. The assessment results that are described and compared in this section are those from the beginning teachers' student teaching and first-year assessments, as well as from their CNTP assessments. The purpose of the comparisons is twofold: (1) to provide some concrete examples of how the CNTP assessments measure the same or different skills as the student teaching and first-year assessments, and, (2) to provide a close-up look at some of the teachers' student teaching and first-year assessment results in order to give a greater sense of the degree to which some of these assessments are or are not coherent and comprehensive. The focus of each comparison is on the assessments' constructs and the teachers' strengths and weaknesses as identified by the assessments.

Before describing the methodology and the comparisons, we would like to add that because our sample of teachers for the comparison is small, any generalizations or conclusions that may be suggested by the results of the comparisons can only be tentative at best. The comparison results can highlight, however, some issues of concern which should be explored further.

Methodology

The Methodology section describes how the innovative assessments and the teachers for which comparisons were made were selected.

Selection of Innovative Assessments

Twelve innovative assessment instruments were pilot tested as part of the CNTP assessment component. These innovative assessments represented a wide variety of assessment approaches -- i.e., high-inference classroom observation, portfolio, semi-structured interview, structured simulation tasks, performance-based assessment center tasks, assessment with videotape stimulus, and assessment with a staff development component. Although all of the assessments were developed sufficiently for the purpose of pilot testing (at least on a small scale), many were still basically in a primitive state of development (especially the scoring systems). For this section of the report, therefore, only those innovative assessments judged to be the most developed were selected for the purpose of comparison with the assessments administered during student teaching and the first year of teaching. These selected assessments were semi-structured interviews, high-inference classroom observations, and portfolios.

Selection of Teachers

As stated above, all 22 teachers who participated in our study had also participated in one of the CNTP pilot tests of innovative assessments. All of these teachers had also

agreed to let FWL collect and examine assessment information from their teacher preparation programs and their school districts. Not all teacher preparation programs and school districts, however, sent the assessment information that FWL requested. Thus, when we looked at those teachers who had participated in the pilot tests of semi-structured interviews, high-inference classroom observations, and portfolios, we only considered selecting those teachers for whom we had complete (or almost complete) assessment information. Of these teachers, we selected two representatives for each assessment method: a teacher who represents the high end of the performance scale for the innovative assessment, and a teacher who represents the lower end. Since there was almost always more than one teacher at each end of the scale, we usually selected, whenever possible, teachers whose assessment results illustrated interesting dilemmas (i.e., highlighted difficulties that could occur from discrepant results).

We would like to state that although some teachers' performance results were at the lower end of the performance scale of some instruments, this does not mean they are poor teachers. The data only suggests that the teachers may have some weaknesses or areas in need of improvement, based on the criteria determined by the assessment.

All of the teachers selected for comparison purposes in this section were given fictionalized names; however, all other data in the comparisons is based on the teachers' actual assessment results and on the teacher interviews.

Assessment Comparisons

The description and discussion of the comparisons of assessment results are presented according to the innovative assessment in which the teacher participated. The assessment results for two teachers who participated in semi-structured interviews are presented first, then the results for the teachers who participated in high-inference classroom observations, and, finally, the results for the teachers who participated in the portfolio assessment.

Semi-structured Interviews

The semi-structured interview approach typically involves a candidate performing a task and then being interviewed about that task. Over the three years of pilot testing, three semi-structured interview assessments were pilot tested: the Semi-Structured Interview in Secondary Mathematics (SSI-SM), the Semi-Structured Interview in Elementary Mathematics (SSI-EM), and the Semi-Structured Interview in Secondary Social Studies (SSI-SS). The SSI-SS has not yet been scored, so we will only look at some of the results of the first two assessments.

Using an interview format, the SSI-SM and SSI-EM were designed to assess a beginning teacher's knowledge in the respective subject areas of secondary and elementary mathematics, exploring a teacher's thought process as he or she makes instructional decisions for students. For the assessments, each candidate performs four or five tasks that pertain to a selected topic in mathematics (e.g., linear equations, fractions) and is asked questions (i.e., is interviewed) after each task. For the SSI-SM, the candidate performed five tasks: Structuring a Unit, Structuring a Lesson, Alternative Mathematical Approaches,

Alternative Pedagogical Approaches, and Evaluating Student Performance. For the SSI-EM, the candidate performed four tasks: Lesson Planning, Topic Sequencing, Instructional Vignettes, and Short Cuts. Although the tasks are not identical, the scoring systems for the SSI-SM and SSI-EM both focus on assessing a candidate's content knowledge, curriculum knowledge, content pedagogy, and knowledge of students. (The scoring system for the SSI-SM, however, was much more developed than that of the SSI-EM.)

If we compare the alternative SSI assessments with those assessments made by the IHEs and LEAs, several things stand out. First, it was clear in our analysis of the IHE and LEA assessments that content knowledge, curriculum knowledge, and content pedagogy are usually only minimally assessed (if at all). The SSI assessments, however, focus heavily on each of these constructs. On the other hand, the IHE and LEA assessments focus strongly on general pedagogical knowledge and performance (e.g., classroom management), which the SSI assessments do not. The CTC assessments such as NTE subject area exams assess content knowledge, but not curriculum knowledge or content pedagogy.

To continue the comparison further, let's look more closely at the performances of the interviewed teachers who took the SSI assessments. Of the 22 teachers we interviewed, four had participated in the SSI-SM or SSI-EM assessments. Of these four teachers, all but one passed the alternative assessment. In fact, one teacher's performance was so good that it was used to define the highest rating point in content knowledge, curriculum knowledge, and content pedagogy. (No candidate received a high rating in knowledge of students, partly because the interview questions did not solicit enough information in this area to make a judgment.)

Let's look now on an individual level at how some of the results from the SSI assessments compare with the results from the assessments experienced by the teachers during student teaching and in their first year of teaching.

Andy – When Knowledge of Subject Matter is Not Enough. We'll begin with Andy, the teacher whose high performance set the standards for the SSI-SM assessment. Andy's knowledge of mathematics was first assessed in his undergraduate years when he was pursuing an English degree. Andy took the NTE math exam and scored very highly. After receiving his English degree, Andy took one year to acquire a math degree. His resulting math GPA was not very high, but this was perhaps a reflection of the fact that he had carried so many units (e.g., at one time, 27) each quarter when pursuing the math degree.

In Andy's two-semester teacher education program, the education coursework was simultaneous with student teaching. His first student teaching assignment consisted of two periods of algebra a day for approximately one quarter and a half. His second student teaching assignment, which lasted for an entire semester, consisted of an English class, a general math class, and later, a pre-algebra class. In each of his two student teaching assignments as a math teacher, Andy received very high praise from his university supervisor and master teacher. However, only in the first student teaching assignment was there any evaluation of his math knowledge. His university supervisor commented that he is "knowledgeable of the math content," and his master teacher said he brings a lot of "knowledge to his teaching." In none of the student teaching evaluations was there any reference to Andy's content pedagogical skills or knowledge of curriculum. Instead, the

evaluations focussed on general pedagogical skills such as his organization and planning of lessons and his classroom management, all of which were highly praised. In the final evaluations made after the second student teaching assignment, his university supervisor declared him "already an effective teacher," and his master teacher remarked that "all elements of good teaching were evident." No weaknesses or areas of improvement were noted.

The glowing assessments of his teaching abilities made by his university supervisor and master teacher could not have prepared Andy for his first year of teaching, which he described as more a matter of "survival" than anything else. Although he had been hired for a job (teaching general math) in the same school where he had done his student teaching, he said he was not prepared for the students he taught, whom he called "difficult." Whereas he had been praised highly for his discipline during student teaching, he now had lots of discipline problems, and this was one of two primary areas noted as in need of improvement on his first-year evaluation by the principal. The other area in which he was assessed as needing improvement was lesson planning (this was again in direct contrast with the high marks he received in this area during student teaching). In particular, the principal encouraged him to present math in "its simplest terms," to "model," to use several instructional approaches, and to check for understanding. The principal also recommended that he create a more active class. Although the principal's evaluation recognized some of Andy's strengths, including "working with high ability kids," knowledge of subject matter was not listed as one of them. In fact, it was Andy's feeling that abundant math knowledge was considered more of a liability than an asset.

Obviously, for Andy, the primary constructs measured by the SSI assessments (i.e., content knowledge, curriculum knowledge, and content pedagogy) were not the primary constructs measured by his IHE and LEA assessments. The IHE assessments focussed primarily on general pedagogical constructs, with only a cursory glance towards content knowledge. The LEA assessments also focussed primarily on general pedagogical constructs, with a fair amount of attention given to content pedagogy. The SSI assessment identified the teacher's content knowledge, curriculum knowledge, and content pedagogy to be strengths. Of those three areas, the IHE assessments identified only his content knowledge as a strength (and made no reference to the other two). The LEA assessment made no direct reference to any of the areas as strengths. Moreover, where the SSI assessment identified content pedagogy as a strength, the LEA assessment identified it as a weakness. What is interesting, however, is that the SSI assessment indicated that Andy's math knowledge is very high, and the LEA assessment noted that Andy works well with high ability students -- a skill requiring, it would seem, high content knowledge. Thus, all three assessments seem to agree that his content knowledge is high, but there is disagreement in how well he applies it.

Beverly -- When Knowledge of Math is Not a High Priority. Our second comparison of SSI results with those from other assessments involves Beverly, one of the elementary teachers who participated in the SSI-EM assessment. Of the four SSI tasks completed (all of which focussed on the topic of fractions), Beverly passed two and failed two, thus warranting an overall rating of "fail." Beverly passed the task on lesson planning, which required her to plan a specific lesson on fractions for a fifth-grade class and then to answer questions about that lesson. Beverly also passed the topic sequencing task on fractions,

although just barely. She was assessed as having little familiarity with several of the topics, including multiplication of fractions and subtraction of mixed numbers with regrouping. Beverly failed the task which required her to respond to a series of hypothetical questions involving students in after-school tutoring sessions which focussed on fractions. She was rated especially low in ability to communicate and motivate students with regard to the specified material. Beverly also failed the task which required an evaluation of computational "shortcuts" in terms of their pedagogical and mathematical soundness. One rater commented that it seemed likely that the "shortcuts" were Beverly's primary methods of instruction.

The results of the SSI assessment indicated that Beverly was judged as being deficient in math skills, especially with regard to the topic of fractions. When we compared the results of the SSI assessment with those made during Beverly's student teaching and first year of teaching, we wondered if any of the latter assessments produced the same results.

We began the comparison by looking at Beverly's first student teaching assignment in which she was in a bilingual primary grade classroom Monday through Friday mornings, for approximately six weeks. The master teacher's evaluation of Beverly's performance indicated that Beverly had had a variety of math teaching experiences during this assignment (e.g., Math Their Way, Patterns, Problem Solving), but there was no assessment made of those experiences, and there was no mention of any experience teaching fractions. Instead, the master teacher's evaluation focussed almost solely on Beverly's personal qualities (e.g., "creative, polite, sincere") and love for children. Although the master teacher evaluated Beverly's performance as "outstanding," no evidence was given to support that evaluation. Also, no weaknesses or areas in need of improvement were cited. For the same assignment, the university supervisor's evaluation described and evaluated Beverly's experiences in four subject areas (two of which focussed on ESL instruction), but math was not one of them. Like the master teacher's evaluation, the university supervisor's evaluation was full of praise for Beverly's performance and made no mention of any weaknesses or areas in need of improvement.

Beverly's second student teaching assignment was in a bilingual intermediate grade classroom for a duration of between four and six weeks. (It is unclear how long this assignment lasted. When interviewed, Beverly said that the assignment was cut short after approximately four weeks because she was asked to substitute for her first master teacher who had become ill. Neither the master teacher nor the university supervisor, however, made mention on their evaluation forms that the assignment was cut short.) On the evaluation form, the master teacher noted that, during the assignment, Beverly taught addition, subtraction, and multiplication of fractions, as well as common denominators, to the students. The master teacher stated that the students showed growth in math "due to [the teacher's] persistence and follow-through in teaching fractions." Interestingly, the university supervisor's evaluation of Beverly's performance made no mention of any instruction in fractions, but noted instead Beverly's "excellent lessons on geometry." The university supervisor again had high praise for Beverly, commending her overall "excellent teaching skills" and describing her as "outstanding." Neither the master teacher nor university supervisor mentioned any weaknesses or areas in need of improvement. (It is also interesting to note that, according to Beverly, the master teacher was absent from the

classroom for much of the assignment because she was working on an extracurricular school activity.)

After completing student teaching, Beverly was hired the following year by the same district as a bilingual intermediate grade teacher. In Beverly's district file, there is one evaluation report made by the principal. (In the interview, Beverly could recall only two assessments made by the principal, and could not recall when the first was conducted.) The principal gave Beverly the highest rating (i.e., "meets objectives") on every one of the form's 21 criteria. In no area was Beverly evaluated as needing improvement, although the principal did recommend that Beverly "continue to explore the many possibilities inherent in instructional strategies, i.e., behavioral and cooperative." When asked to recall if there had been any weaknesses mentioned on the evaluation reports, Beverly replied, "I called on one girl too often who was aggressive." (This "weakness" was not listed on the evaluation report.)

Thus, although Beverly had failed the SSI-EM assessment, indicating serious math deficiencies with regard to the topic of fractions, she "passed" all of the assessments during student teaching and in the first year of teaching, and there was never the slightest indication in any of these assessment results of a deficiency in math with regard to any topic. The discrepancy in results suggests that the SSI-EM assessment covers different constructs than those assessed during Beverly's student teaching assignments and the first year of teaching. The SSI-EM assessment assessed Beverly's skills in-depth in one area, while the other assessments seemed to assess her skills much more broadly and across a multitude of areas. The difference in results from the various assessments is also noteworthy for the following reason: the results of the SSI-EM assessment seem to indicate that Beverly has some severe weaknesses but no strengths, while the results of the other assessments seem to indicate that Beverly has only strengths and no weaknesses. Neither set of results seem appropriate for a beginning teacher.

Classroom Observations

Although the classroom observation approach to assessment is the approach that has been traditionally used by university and district personnel to assess a beginning teacher's ability to teach, most classroom observation instruments have been little more than observation checklists. In our three years of pilot testing, we had the opportunity to pilot test two innovative classroom observation instruments that differed from more traditional instruments in that they were highly inferential in nature, and they required substantial training on the part of the observers. In addition, one of the instruments was designed specifically to assess a beginning teacher's content pedagogical skills in a particular subject matter. The two instruments pilot tested were the Connecticut Competency Instrument (CCI) and the Laboratory Science Assessment.

The CCI was designed to assess a teacher's general pedagogical skills at any grade level. Developed by Connecticut for use in their credentialing process, it had already been field tested and was highly developed. In contrast, the Laboratory Science Assessment was newly developed, and, for all practical purposes, had never been pilot tested. After our pilot test of the instrument, we determined that both the instrument and its scoring system

needed many revisions. Hence, for this report, we will only look at the performance results of the teachers who participated in the CCI pilot test.

The CCI assessment consists of a pre-assessment information form filled out by the teacher before the observation, a pre-observation conference, a 45-60 minute observation (during which time a trained observer scripts the lesson as accurately as possible), and a post-observation conference. After conducting the assessment, the observer utilizes the various sources of information (e.g., conferences, observation) to rate the teacher's performance as "acceptable" or "unacceptable" along 10 specifically defined indicators which are grouped under three categories: Management of the Classroom Environment, Instruction, and Assessment. In support of each "acceptable" or "unacceptable" rating, the observer is required to provide in writing very specific positive and/or negative evidence from the conferences and observation as it applies.

Of the 22 teachers we interviewed, three participated in the CCI pilot test. Of those three, we only have complete assessment information for one. We will look at the CCI results for that one teacher and see how they compare with the results of the other assessments of the teacher made during student teaching and the first year of teaching. We will also look at a second teacher's CCI results, but, due to the lack of information, will be able to make only a limited comparison with the teacher's other assessments.

Colleen – Every Observation Tells a Similar Story. Our first teacher, Colleen, participated in the CCI pilot test when she was a teacher of a bilingual primary grade. She was observed teaching a math lesson. Her CCI results show that not only did she pass all 10 indicators, but there was almost no negative evidence written for any of the indicators. The only negative evidence to be found was under the indicator, Lesson Content, where Colleen was faulted for having misspelled a word, and under the indicator, Monitoring and Adjusting, where the observer noted that Colleen responded incorrectly to a student error. For the rest of the 10 indicators, there was only ample and positive evidence of Colleen's general pedagogical skills as they corresponded to each of the indicators.

The assessments made of Colleen's general pedagogical skills during her two student teaching assignments -- both of which were in the primary grades and in bilingual classes-- were also positive. In Colleen's three-semester teacher education program, student teaching began in the second semester. Her first assignment lasted 10 weeks, and she student taught Monday through Friday in the mornings only. Her second assignment also lasted 10 weeks, but she was at the school site daily for the entire day. After both assignments, neither the master teacher nor university supervisor identified any weaknesses or areas in need of improvement (although, after the first assignment, the master teacher seemed to hint at a weakness when she commented that Colleen's motivation techniques had improved as had her confidence in dealing with children who were discipline problems.) Instead, the master teachers and university supervisor had only positive things to say about Colleen's performance in planning, teaching, assessing pupil progress, relating to pupils and adults, as well as about her personal attributes.

The evaluation made of Colleen's first year of teaching was also very positive. The principal commented positively on her performance in six areas: planning, instructional strategies, evaluating student progress, classroom management, professional growth, and

personal responsibilities. One of the comments made was that Colleen's first year had proven to be "outstanding." No weaknesses or areas in need of improvement were mentioned.

Overall, the various classroom observation assessments that had been made of Colleen's general pedagogical skills seemed to reach the same conclusion: Colleen was competent in the area of general pedagogy. Moreover, all the assessments seemed to cover the same general pedagogical constructs. That is, all the assessments (i.e., the CCI assessment, the student teaching assessments and the first-year assessments) assessed Colleen's skills in planning, using instructional strategies, evaluating student progress, and classroom management. Perhaps the biggest difference among the assessments was that the CCI assessment defined the constructs as they pertained to one lesson, while the other assessments evaluated the constructs as they pertained to a time period of eight weeks to one school year. For example, as shown by Figure A.1, the CCI assessment's evaluation of Colleen's ability to monitor student progress focussed on her ability to monitor student understanding during the lesson and to adjust instruction when needed. In contrast, the first-year assessment focussed on evaluation skills not necessarily pertaining to the teaching of one lesson (e.g., communicates student progress to students and parents). Thus, although the same general constructs were evaluated by all of Colleen's assessments, the pictures painted by the various assessments of Colleen's pedagogical skills were not by any means identical.

Deanna – When Subject Matter Might Make a Difference. In contrast to the above teacher, Deanna, who was also observed teaching a primary grade class, failed the CCI assessment. Deanna, who was assessed while conducting a reading lesson, received an unacceptable rating in the following areas: learning environment, student engagement, structure for learning, lesson development, questioning, communication, and monitoring/adjusting. Although the CCI evaluation form included positive evidence in each of the above areas, the preponderance of evidence in each area was negative, thus warranting an overall rating of unacceptable.

During her two-semester teacher education program, Deanna completed two student teaching assignments, the first eight weeks in length, the second, 10 weeks. In both assignments, Deanna was at the school the entire day, Monday through Friday. Of Deanna's two student teaching assignments, we only have one official evaluation from the first assignment which was completed in a primary grade. This evaluation appears to be completed by Deanna's university supervisor, and, based on the supervisor's description, seems to be of a science lesson. According to the evaluation form, there were five general areas to be assessed by the observation: planning and evaluation, classroom organization and structuring, teaching strategies, classroom interpersonal climate, and personal/professional. As none of these areas are defined in any way on the form, it is difficult to compare them with the CCI assessment categories, but on the surface, it appears that the student teaching assessment was designed to cover some of the same general areas as the CCI assessment. The reality of the assessment, however, was different from the design. Of the five general areas listed on the form, the supervisor's assessment only addressed Deanna's planning skills and teaching strategies, both of which were positively evaluated. The supervisor also added that Deanna's knowledge of subject matter was a

FIGURE A.1

ABILITY TO ASSESS STUDENT PROGRESS

(As Defined by Criteria from Colleen's Various Assessments)

CCI Assessment

- Monitors the level of student understanding at appropriate points during the lesson.
- Adjusts teaching when necessary (e.g., when monitoring indicates students are misunderstanding).

Student Teaching Assessments

Evaluation forms indicate that the definition is open to interpretation.

First Year Assessment

- Uses evaluation of student outcomes to measure student progress.
- Communicates student progress to students and parents.

strong point "that will always place you in the realm of successful teachers." The supervisor noted no weaknesses or areas in need of improvement.

Although our records contain no official evaluation from the second student teaching assignment, which was conducted in an upper elementary grade, we do have a letter of reference on file written by the school principal. In this letter (which was longer than the formal student teaching evaluation from the first assignment), the principal describes Deanna as a "person with a warm personality who projects a positive attitude." This evaluation is in direct contrast with the CCI assessment in which the observer specifically noted Deanna's lack of warmth and enthusiasm. The principal also described in detail a math lesson he had observed, commenting that "most of the students remained on task." Again in contrast, during the CCI observation of a reading lesson the observer noted that, on average, only 76% of the students were on task, thus warranting a rating of unacceptable in the area of student engagement. Although overall the letter seems to be a positive recommendation, and the principal never directly mentions any weaknesses, the principal does remark at the letter's end that "at this time" the primary grades would be most appropriate for Deanna. The principal gives no indication why this is so, but possibly the reason may be found in our interview with Deanna.

When interviewed, Deanna reported that her second student teaching assignment was in an ethnically diverse classroom and that she was assigned to this classroom because the master teacher was having trouble and needed help with the many students in the class who had emotional problems. Deanna said that her university supervisor felt she was too rigid with the students and was too hard on them. Deanna's perception was that these students needed a teacher who was in control. She felt the supervisor's opinion of her as lacking empathy with minority students was unfair and untrue, and added that the master teacher did not share the supervisor's opinion.

Whether the supervisor ever wrote an evaluation of Deanna's skills in working with diverse students is unknown. Such an evaluation, if it does exist, was definitely not part of Deanna's district file. Perhaps the principal was alluding to this problem in his statement, but, because of the absence of assessment information on file, it is impossible to know for sure.

The CCI assessment makes no mention of the ethnic make-up of the class observed, and none of the assessment indicators are specially designed to assess a teacher's skills in working with diverse students. Thus, although Deanna's skills in working with diverse students may have been in need of improvement, neither the CCI assessment nor the student teaching assessments in the district file address this area.

In the interview for her first job, Deanna said she brought up the problem she had with the upper grade students, but said the interview people were not too concerned about management issues. She described the school atmosphere as relaxed and noted that most of the student population was from a high socio-economic class. The interviewers were much more interested in her math and science background, a rarity among elementary teachers.

Looking at the district file, there does not seem to be any evaluation of Deanna's first year of teaching. Thus, no comments or comparisons can be made about Deanna's first-year assessment. There are, however, evaluations of Deanna's second and third year of teaching. Each evaluation, completed by the principal, rates Deanna as meeting district standards in the following four areas: student progress, professional competence, control and learning environment, and required duties and professional responsibilities. The two evaluations are almost identical, with very brief comments offered regarding the students' academic progress, Deanna's attendance at workshops, the orderliness of the classroom, and the fact that all reports are done on time. No weaknesses or areas in need of improvement are mentioned.

In comparison with the CCI assessment, the only area in which there seems to be some overlap is in the assessment of the teacher's control and learning environment. However, the district evaluations comment favorably on Deanna's skills in this area, while the CCI assessment (conducted in Deanna's first year of teaching) does not. The district evaluations make no assessment of the other areas assessed by the CCI such as lesson content, lesson development, communication, etc. Another area of difference is that the district's year-two evaluation compliments Deanna on her science and math instruction, an area not noted in the CCI assessment which focussed on a reading lesson.

When looking at Deanna's assessments overall, several things stand out: First, the CCI assessment was a much more detailed and thorough assessment of her teaching skills than the student teaching or district assessments. Second, based on references made in the student teaching and district assessments, it seems likely that a CCI assessment conducted of a math or science lesson may have produced better results than did the assessment of Deanna's reading lesson. (At the very least, Deanna probably would have been judged as more enthusiastic.) Third, based on Deanna's interview responses and a comment made in a letter of reference, Deanna's skills in working with diverse students may be in need of improvement, but none of the assessments focussed on this area. And fourth, although the CCI assessment evaluated Deanna in her first year of teaching and found serious deficiencies in her teaching skills, her employing district has evaluated her performance over the last several years and found her skills to meet district standards, with no deficiencies noted.

Portfolios

Portfolio assessments require teachers to demonstrate their teaching skills through a collection of artifacts and/or documentation of activities representing their teaching during a particular period of time (e.g., a school year) or of a particular unit of instruction. The portfolio assessment pilot tested by the CNTP project required secondary English teachers to both collect artifacts of their teaching and document their teaching activities for a particular unit of instruction (their choice) in English or the language arts. The assessment was designed to evaluate a teacher's skills in three areas: planning and implementing a unit, responding to student work, and reflecting upon the experience in teaching a unit to gain insight for further teaching.

For the portfolio assessment, each participating teacher planned and conducted a three- to six-week teaching unit in which the classroom activities were unified by a single

focus (e.g., a novel, a particular genre, a set of skills). To document the teaching activities, each teacher compiled a portfolio which consisted of the following components: an outline of a unit plan, a weekly log, all materials and assignments given to the students, samples of student work with teacher responses, student evaluations (of the unit or of one major activity), and a reflective essay.

Scoring of the portfolio assessment was complex and varied. Each portfolio was independently evaluated by two trained scorers, each of whom used a scoring response form divided into six parts, with each part representing a different skill or teaching competency. The six skills or competencies assessed were as follows: planning abilities, unit design, portfolio presentation, general pedagogical abilities, subject-specific pedagogical abilities, and reflective ability. For each part, the teacher's performance was first evaluated according to specific criteria listed for the skill/competency, and then an overall rating was awarded. The criteria and overall ratings were made using a three- and/or four-point scale, and all ratings were made in a holistic manner and were not interdependent.

Four of the 22 teachers interviewed participated in the portfolio assessment. One of the four teachers was given passing ratings on all six portfolio parts; the other three teachers passed either one or none of the portfolio parts. We shall first look at the assessment results for the teacher who passed the portfolio assessment, and then at the results for a teacher who did not.

Elizabeth -- Outstanding from the Start. Elizabeth, the teacher who passed the portfolio assessment, had been teaching English for two and a half years when she completed her portfolio. For the assessment, Elizabeth planned a five-week unit on a classical novel and then taught the unit to a class of predominantly Caucasian students (75%) in a suburban high school. For each of the six overall scores, the two scorers who evaluated her portfolio gave her either a "3" or "4" rating (3 = some strength in this area; 4 = definite strength in this area). Both scorers agreed that she showed some strength (i.e., a "3" rating) in the area of planning abilities, and definite strength (i.e., a "4" rating) in her unit design. Regarding the latter, both scorers were particularly impressed with Elizabeth's many and wide variety of activities, including cooperative endeavors and activities which encouraged individual reflection on the book's meaning.

In the other four areas, one scorer consistently gave Elizabeth ratings of "4" (i.e., has definite strength in these areas), and the other scorer consistently gave her ratings of "3" (i.e., has some strength in these areas). In the overall evaluation of her general pedagogical abilities, for example, one scorer gave her a "3" and the other a "4," and both scorers rated her as having some or definite strength with regard to, among other things, flexibility and adaptability, understanding student attitudes and feelings, and choice of appropriate evaluation methods.

It is somewhat difficult to compare Elizabeth's portfolio assessment results with the evaluations made of her teaching skills during student teaching because Elizabeth's original plans were to become an elementary teacher. After receiving her B.A. in English, Elizabeth attended a special UC program whereby she qualified for both a master's degree and a multiple-subjects credential. Elizabeth's first student teaching assignment was in a university lab school where she student taught in a primary grade class every morning,

Monday through Friday, for approximately 11 weeks. On her final evaluation, she received mostly "outstanding" ratings and some "good" ratings (the two highest ratings on a five-point scale) on 21 criteria. Among other things, Elizabeth's instructional planning skills were rated as outstanding, as was her knowledge of content in the subject(s) taught and the procedures she used to assess student progress. Elizabeth's master teacher also commented positively on her planning skills and content knowledge in the written commentary which was part of the evaluation. No weaknesses or areas in need of improvement were mentioned.

Elizabeth's second student teaching assignment was at a different school located in a wealthy urban area, but which had close to 40% of the student population bussed in from the inner city. She again student taught in a primary grade class every morning, Monday through Friday, for approximately 11 weeks. On her final evaluation for her second assignment, all 21 criteria were given ratings of "outstanding" or "outstanding plus" (the latter rating being a creation of the master teacher). In the written commentary, the master teacher described Elizabeth as a "superior student teacher," and praised her language arts unit for, among other things, her clear objectives and variety of activities. The master teacher also noted Elizabeth's "excellent background and knowledge and ideas," her flexibility, and her "effective" evaluation methods. Again, no weaknesses or areas in need of improvement were mentioned.

It was during the second student teaching assignment that a former teacher recommended Elizabeth for a substitute position as a high school English teacher the following year. Elizabeth accepted the job and over the course of the year acquired a secondary English credential.

In Elizabeth's first year of teaching secondary English classes, she was evaluated twice, and each time received the highest rating (i.e., "satisfactory") on each of the district's 13 listed performance criteria. Her first evaluation particularly commended her for her evaluation techniques and her integration of activities. It also noted one area of weakness: a tendency to occupy center stage, thus limiting student involvement. Her second evaluation commended Elizabeth for the progress she had made in several areas, including student involvement. In particular, she was commended for moving away from the practice of telling students the meaning of passages and books, and instead using activities that encouraged the students to think for themselves. No new weaknesses or areas in need of improvement were mentioned.

Overall, there seemed to be strong overlap among Elizabeth's assessments in the areas of planning, unit design, and evaluation. In addition, each of these areas was evaluated as an area of strength for Elizabeth by all of the assessments. There were, however, some major differences among the assessments. For example, some skill areas covered by the portfolio assessment, but not the student teaching and first-year assessments, were reflection and subject-specific pedagogy. Some skill areas covered by the student teaching assessments and first-year assessments but not the portfolio assessment were classroom management, classroom environment, professional responsibility and interpersonal relationships. Another major difference between the portfolio assessment and the other assessments is that the portfolio assessment provided evidence to support the evaluations made while the other assessments did not.

It might also be worthy to note that judging by Elizabeth's first-year assessments, it is possible that the unit design skills which were so highly evaluated by the portfolio assessment (e.g., the inclusion in her unit of activities that encourage individual student reflection on a book's meaning) were greatly enhanced in Elizabeth's first year of teaching under the guidance of her principal. Elizabeth seemed to receive specific feedback in her first year of teaching with regard to the integration of student-centered activities. Without this feedback, it is possible that Elizabeth's performance on the portfolio assessment would not have been as strong.

Grace -- A Teacher in Progress. The second portfolio teacher we will look at is Grace. Like Elizabeth, Grace had already completed one year of teaching before participating in the portfolio assessment. For the assessment, Grace planned a four-week unit based on a short novel that she selected from a district-approved recreational/motivational reading list. Grace taught the unit to a ninth grade remedial English class, composed primarily of Hispanic students. In her outline of the unit plan, Grace said her primary reason for her choice of the book was that it was easy to read and follow, and that her major objective was for as many students as possible to read the entire novel with a high level of recall.

Of the six competency areas scored, Grace did not receive a single passing rating. Moreover, for all but two of the areas, both scorers gave Grace a rating of "1" (i.e., serious weaknesses in this area). Only in the evaluation of planning abilities did both scorers give her a higher rating of "2" (i.e., minor weaknesses in this area). Ratings for her portfolio presentation were mixed, with one scorer giving her a "1" and the other a "2."

The scorers' comments about Grace's portfolio explain why they gave her such low ratings. For example, her unit design was given a "1" rating because her unit activities were primarily composed of oral/silent reading and the answering of teacher-generated recall questions. The plan left little room for student input (except to answer recall questions) and showed no creativity. Her pedagogical abilities were given a "1" rating because her strategies were limited and her understanding of students was minimal. She was described as tightly controlling the class through a "rigid lesson plan and rigid quizzes." She was also faulted for not responding to student failure by changing the unit or attempting to assist the students. Her subject-specific abilities were given a "1" rating largely because of the limited writing opportunities she provided students (e.g., responding to multiple-choice and fill-in-the-blank tests), the limited speaking opportunities she provided (e.g., oral reading and class discussion "as needed"), her lack of creative assignments that promoted thoughtful responses to literature, and her limited awareness of the language needs of Hispanic students. Finally, her reflective abilities were given a "1" rating because in her reflective essay she never evaluated her lack of success (as indicated by negative student evaluations of the unit) or the students' lack of success. Also, although Grace wrote that she felt good about her teaching, her essay revealed a negative bias towards her students and their culture.

Grace's discomfort with students who are different than she may be a result, in part, of her lack of exposure to such students. Grace attended school outside of California in a state where non-Caucasians are truly a minority. After receiving a B.A. in English, Grace attended a teacher preparation program where she completed two student teaching

assignments (conducted simultaneously) in approximately a six-week period. Both assignments were at the same high school but in different grade levels. One assignment consisted of two low-ability classes and the other consisted of two advanced classes. A different master teacher was in charge of each assignment.

The information sent to us from Grace's teacher preparation program file was extensive. Included were not only her final evaluation reports, but also her midterm evaluation reports and copies of other formal observation reports made by her master teachers. Each of these latter reports consisted of an abbreviated transcript (4-5 pages) of an observed lesson, some summary suggestions, and a formal evaluation based on the observation. All of the different evaluations used a three-point scale ("1" = needs additional work, "2" = acceptable, "3" = definite strength) to rate her performance along numerous criteria.

The midterm evaluation forms were slightly different than the final evaluation forms, although both covered the same three areas: Personal Qualities, Classroom Management, and Teaching Techniques. Grace's midterm evaluations by both master teachers were primarily composed of "2" ratings. Both master teachers rated Grace as needing improvement in anticipating and preventing discipline problems and allowing for differences in individualizing instruction. The master teacher of the two low-ability classes described her instructional approach as conservative, stating that "questioning and reading aloud is the primary focus of class time." The master teacher of the two advanced classes rated Grace a "1-2" on "knows subject matter well" and commented that Grace did not seem comfortable with the advanced classes and the classes seemed to lack confidence in her.

On her final evaluation, Grace's ratings had improved to the point where the master teacher of the low-ability classes gave her an equal number of "2"s and "3"s. This teacher listed Grace's strengths as patience in discipline and empathy for students. The areas in need of improvement were developing effective lessons and the presentation of lessons. The master teacher of the advanced classes gave Grace mostly "2"s, with some "3"s. The strengths listed were initiative and motivation and her questioning skills. The specific areas listed as in need of improvement were the type of feedback she gave to students (i.e., she needed to be more responsive and accepting of student answers) and discipline. Regarding the latter, the master teacher commented that Grace was often "overly rigid." She also commented that Grace appeared to be more comfortable with average/lower level students.

Although Grace sent her application for a teaching position to numerous districts in her state, her first job was in an inner city California high school. She had answered an ad placed in the local paper by the district and was subsequently interviewed by the district superintendent and associate superintendent. Grace recalls that they asked her many questions about her discipline and general teaching styles and about her skills in working with diverse student groups, especially minorities and low-level students. Grace remembers being told repeatedly that the position was for "basic" classes. Grace thought the term somehow referred to grammar. She later found out basic meant remedial.

Like her file from her teacher preparation program, Grace's district file included much information pertaining to the evaluation of her teaching skills. In addition to two summary evaluation reports (one completed at mid-year and the other at the end of the

year), a total of six observation reports were also included. Each observation report, completed by the assistant principal, included a description of the instructional situation, indicators of student progress and control, and specific suggestions for needed improvement. The reports indicate that there were many areas in need of improvement and that the assistant principal gave Grace many specific suggestions over the year related to general pedagogical techniques such as initiation, closure, questioning, student involvement, student discipline, and choice of lesson activities.

Grace's summary evaluation reports, also completed by the assistant principal, consisted of 30 criteria distributed under five evaluation areas: Instructional Skills, Learning Environment and Student Control, Duties and Responsibilities, Health and Personal Characteristics, and Professional Relationships and Professional Development. The two reports were almost identical with all criteria marked "Satisfactory," the highest of three possible ratings. The mid-year report also complimented Grace's "positive attitude" and indicated that she had made progress in her use of lesson design techniques and assertive discipline rules. The final report also referred to Grace's trying different teaching techniques and to her growth in class management techniques.

By looking just at Grace's student teaching and first-year assessments, the picture seems to be of a beginning teacher who is just starting to develop her teaching skills and presence in the classroom. Grace appears to be willing to work at improving her skills, and her evaluations indicate that she does improve over time. When one looks at the portfolio assessment, however, conducted in her second year of teaching, the picture is not so clear. Despite all the earlier recommendations made during student teaching and the first year of teaching, Grace seems to have reverted back to her original teaching style which was marked by a rigidity in control and activities reflecting minimal variety, creativity, and student involvement. Here is a case of the alternative assessment revealing problems that have always been there but were supposed to have been improved. Unlike the student teaching and first-year assessments that indicate that Grace is making satisfactory progress, the alternative assessment reveals that some serious weaknesses still remain.

For the most part, there is a fair amount of overlap in the pedagogical constructs assessed by the portfolio assessment and the student teaching and first-year assessments. The portfolio assessment, however, has more of a focus on subject-specific pedagogical skills than do the other assessments. That is, as an English teacher, the portfolio assessment expects Grace to have specific pedagogical skills related to the areas of composition, literature, oral and language performance. None of these areas are specifically evaluated by the student teaching and first-year assessments.

With regard to the identification of strengths and weaknesses, the portfolio assessment does not identify the same strengths as the student teaching assessments (e.g., patience, motivation). Instead, it seems to suggest that Grace has no strengths. It does, however, identify many of the same weaknesses as the student teaching and first-year assessments (i.e., the first-year observation reports). Moreover, the fact that the same weaknesses are identified at a later stage of development as at an earlier stage, lends support to the portfolio assessment's labeling Grace's weaknesses as serious.

One more thing to consider. It was no accident that Grace was recruited from out-of-state. Grace's district has high teacher turnover, and, for a number of reasons largely beyond its control, severe difficulty in attracting graduates of nearby teacher preparation programs. Grace may have serious weaknesses in subject-specific pedagogy as identified by the portfolio assessment, but her district assessment, while identifying many of the same weaknesses, views Grace as progressing satisfactorily. Grace's district has an extremely low resource base, and assistance to teachers who are perceived as performing satisfactorily is unlikely to be able to successfully compete for budget allocations. Based on interviews with similar districts, Grace's district may evaluate her, at least in part, in comparison with the general experience with the population from which replacements might be recruited. Grace "failed" the portfolio assessment but she passed the district assessment, and, for the district, that is probably sufficient.

Conclusions

The assessment results for each teacher described above are briefly summarized in Chart A.1. As can be seen under "Constructs Assessed," the focus of the CNTP innovative assessments generally differ from those of the IHE and LEA assessments in that content pedagogy is more often the focus of the CNTP innovative assessments (with the exception of the CCI). It is less common for the IHE and LEA assessments to focus on content pedagogy; general pedagogy, often accompanied by personal attributes and professional responsibilities, is more often the focus. The aspects of teaching that were absent from the CNTP innovative assessments were often noteworthy to districts, e.g., Beverly's persistence with students, Deanna's strength in math and science, and Grace's ability to control her classroom.

As can be seen in the narrative description of assessment results, often the CNTP innovative assessments focus on teaching skills similar to those forming the focus of the IHE and LEA assessments. However, the CNTP innovative assessments evaluate the teacher's performance from the perspective of content pedagogy (e.g., are the students actively engaged in activities which are likely to increase their understanding of the content -- evaluated by someone familiar with both the content and content pedagogy), while the LEA assessments and many of the IHE assessments evaluate the performance from a more general perspective (e.g., are the students actively engaged in activities which are matched to the teacher's lesson objectives -- as evaluated by someone who is not necessarily knowledgeable about the content, and especially, content pedagogy). The IHE assessments seem more variable, depending on the individual assessor's interpretation of the evaluation criteria.

The assessment results show the importance of assessing application skills. Several teachers (Andy, Beverly, Deanna, and Grace) had quite different evaluations from their district than from the CNTP innovative assessments. Only Elizabeth's and Grace's evaluations detected the same strengths or weaknesses as the CNTP performance evaluation. The semi-structured interviews rely on teacher descriptions of hypothetical lessons. The CCI and the portfolio assessment directly assess application of teaching skills. However, the CCI does not address content pedagogy well, and only addresses a single lesson at one point in time. The portfolio examines application over time, although it

CHART A.1

COMPARISON OF ASSESSMENT EVALUATIONS

Teacher: Andy

Assessment	Constructs Assessed	Positive Evaluation and/or Identified Strengths	Identified Weaknesses	Other Comments
SSI-SM (Focus: Linear Equations)	• Content Knowledge • Curriculum Knowledge • Content Pedagogy • Knowledge of Students	Positive Evaluation Strengths = all areas but knowledge of students	None	High performance set scoring standards for SSI-SM assessment
Student Teaching (2 18-week assignment; 1st in high school, 2nd in middle school)	• Content Knowledge • General Pedagogy • Professional Responsibilities	Positive Evaluation Strengths = organization, planning, classroom management	None	"Already an effective teacher" (US), and "all elements of good teaching evident" (MT), after 2nd assignment
District (Urban middle school)	• General Pedagogy • Content Pedagogy • Professional Responsibilites	Less than Positive Evaluation Strength = working with high-ability students	• Discipline • Lesson Planning • Content Pedagogy	Advised by PR to present math in simplest terms to create a more active class

CHART A.1 (cont'd)

COMPARISON OF ASSESSMENT EVALUATIONS

Teacher: Beverly

Assessment	Constructs Assessed	Positive Evaluation and/or Identified Strengths	Identified Weaknesses	Other Comments
SSI-EM (Focus: Fractions)	• Content Knowledge • Curriculum Knowledge • Content Pedagogy • Knowledge of Students	Not a positive evaluation	• Content Knowledge • Content Pedagogy	Passed lesson planning task; barely passed topic sequencing task
Student Teaching (2 6-week assignments in bilingual classrooms)	• Personal Qualities • General Pedagogy	Positive Evaluation Strength = "excellent lesson in geometry" (US)	None	"Outstanding" (US, after 2nd assignment); "excellent teaching skills" (US)
District (Bilingual classroom)	21 separate criteria	Positive Evaluation	None	Advised by PR to explore different instructional strategies

CHART A.1 (cont'd)

COMPARISON OF ASSESSMENT EVALUATIONS

Colleen

Assessment	Constructs Assessed	Positive Evaluation and/or Identified Strengths	Identified Weaknesses	Other Comments
CCI (Observation of math lesson)	• General Pedagogy	Positive Evaluation	None	
Student Teaching (2 10-week assignments in bilingual, primary grade classrooms)	• General Pedagogy • Relating to Pupils and Adults • Personal Attributes	Positive Evaluation	None	
District (Bilingual, primary grade classroom)	• General Pedagogy • Professional Growth • Personal Responsibility	Positive Evaluation	None	"Outstanding" first year (PR)

CHART A.1 (cont'd)

COMPARISON OF ASSESSMENT EVALUATIONS

Teacher: Deanna

Assessment	Constructs Assessed	Positive Evaluation and/or Identified Strengths	Identified Weaknesses	Other Comments
CCI (Observation of reading lesson)	• General Pedagogy (10 areas)	Not a Positive Evaluation	General Pedagogy (7 areas)	
Student Teaching (Observation of science lesson in first assignment)	• General Pedagogy • Personal/ Professional	Positive Evaluation Strengths = science and music	None	No evaluation available for 2nd assignment
District (2nd and 3rd year assessments)	• Student Progress • Professional Competence • Control & Learning • Professional Responsibilities	Positive Evaluation	None	

CHART A.1 (cont'd)

COMPARISON OF ASSESSMENT EVALUATIONS

Teacher: Elizabeth

Assessment	Constructs Assessed	Positive Evaluation and/or Identified Strengths	Identified Weaknesses	Other Comments
Portfolio (Classical novel taught to predominantly Caucasian students in suburban high school)	• Planning Abilities • Unit Design • Portfolio Presentation • General Pedagogy • Content Pedagogy • Reflective Ability	Positive Evaluation Strength = Unit Design	None	Scorers impressed with many and wide variety of activities
Student Teaching (2 11-week assignments in elementary schools)	21 separate criteria	Positive Evaluation Strengths = planning skills, content knowledge, evaluation methods	None	2nd assignment: all 21 criteria rated as "outstanding" or "outstanding plus"; "superior student teacher" (MT)
District (High school English classes)	13 separate criteria	Positive Evaluation Strengths = evaluation methods, integration of activities	First Evaluation: Need for more student involvement Second Evaluation: None	

CHART A.1 (cont'd)

COMPARISON OF ASSESSMENT EVALUATIONS

Teacher: Grace

Assessment	Constructs Assessed	Positive Evaluation and/or Identified Strengths	Identified Weaknesses	Other Comments
Portfolio (Short novel taught to predominantly Hispanic students in urban high school)	• Planning Abilities • Unit Design • Portfolio Presentation • General Pedagogy • Content Pedagogy • Reflective Ability	Not a positive evaluation	All areas	Serious pedagogical weaknesses; appears negatively biased towards minority students
Student Teaching (2 simultaneous assignments in a 6-week period)	• General Pedagogy • Personal Qualities	Positive Final Evaluations Strengths = Patience, motivation, questioning skills	• Lesson Development & Presentation • Discipline • Feedback to Students	Seemed more comfortable with lower-level students; tended to be "overly rigid" (MT)
District (Urban high school English classes)	• General Pedagogy • Professional Duties • Personal Qualities • Professional Relationships & Development	Positive Summary Evaluation	None in summary evaluations	Observation reports revealed many pedagogical areas in need of improvement

strongly relies on a teacher's writing and organization skills, which is still one step removed from application.

Generally, the CNTP innovative assessments (with the portfolio being an exception) are confined to an in-depth focus on teaching skills at a particular point in time, while the student teaching and district assessments can and usually do evaluate a teacher's abilities over a longer time period. For example, the Semi-Structured Interviews and the CCI both focus primarily on activities relating to lessons or a short series of lessons, while the IHE and LEA assessments can focus on more long-term goals, such as use of student evaluations to guide instruction. Paradoxically, however, the use of trained subject-matter teachers by the CNTP innovative assessments which assess content pedagogy perhaps better position these assessments to diagnose potential problems that will not appear until the long term, e.g., sequencing of instructional activities to achieve a long-term learning goal. The LEA use of evaluators from the teacher's own context facilitates the detection of problems that will appear in the short run, e.g., student engagement or classroom management. But evaluators who lack the appropriate expertise in content pedagogy are unlikely to identify more long-term problems related to student evaluation and the curriculum.

The more long-term view characteristic of IHE and LEA assessments should provide opportunities for evaluation and feedback with respect to a teacher's growth. However, the lack of a developmental perspective of teaching described in this and earlier chapters is illustrated by the descriptions of identified strengths and weaknesses. At the end of student teaching, Andy is described as "already an effective teacher" and Beverly is described as possessing "excellent teaching skills". Other teachers not profiled were also described as having all the requisite skills needed for teaching. This overly positive feedback need not be the case. For example, take the instance of Andy, who received very positive evaluations of his classroom management during student teaching and very negative evaluations during his first year. Classroom management is an instance where it is difficult to fully evaluate a student teacher's skills, because the class is still under the ultimate control of the master teacher. Andy could have been told during student teaching that his use of classroom management techniques for *that class* was good. For instance, Elizabeth is described as a "superior student teacher," which does not imply that she has fully mastered her craft.

While summative evaluations of student teaching will necessarily be overwhelmingly positive for those who pass, the evaluations received send an important and powerful message to beginning teachers. After receiving student teaching evaluations that imply that they have little more to learn, beginning teachers must be needlessly frustrated as they find that this is not the case. As the chart illustrates, weaknesses *are* identified for a few teachers at a few points in time. For the most part, however, the written documentation examined confirms the teachers' reports that little feedback is given through the IHE and LEA summative assessments. The CNTP innovative assessments in their present form are little better. The developmental perspective is usually reflected in the choice of criteria and the setting of standards at levels which are commensurate with beginning teacher abilities. However, the assessment results again do not indicate that further development is needed for successful beginning teachers, because weaknesses that are considered to be typical of beginning teachers are not mentioned. The communication of the developmental perspective to beginning teachers by the CNTP innovative assessments depends upon as yet undeveloped descriptive materials which clearly articulate a developmental perspective

against which the assessment results are interpreted. However, CNTP alternative assessments *are* likely to identify some weaknesses and areas in need of improvement, whereas IHE summative assessments are not. LEA assessments are likely to identify major weaknesses in teaching skills; minor weaknesses are likely to go unemphasized at best, unnoticed at worst.

In examining the assessment documents, we found that documents varied in their ability to convey evaluative information. Assessment forms which consisted of a checklist of rating categories by skills which were not accompanied by additional documentation gave only the most general impression of a teacher's skills. Supporting documentation (e.g., observation notes) gave meaning to and supported the summary comments. Definition of the skill and rating categories (which was rarely the case) improved the ability to interpret the information. Assessments which provided a lot of detailed evidence to support the ratings were the most helpful for interpreting the evaluation information, and seem most helpful for purposes of development or improvement. These assessments not only included most of the CNTP innovative assessments but student teaching and first-year assessments which included observation transcripts. Thus assessments with descriptive evidence seem to have high potential for use by beginning teachers, employers, and support providers who can be best served by an accurate picture of a teacher's skills.

Interpretation of assessment results was particularly problematic for elementary teachers, who constituted half of the teachers profiled. Elementary teachers teach multiple subjects, and cannot be expected to be equally expert in each. The CNTP innovative assessments focussed on the teaching of a particular subject. There were hints for two of our three elementary teachers (Deanna and Beverly) that their strengths lay in subjects other than that which was the focus of the alternative assessment. Neither the CNTP innovative assessments in their present conceptualization nor the IHE and LEA assessments seem to tackle the problematic issue of how to evaluate elementary teachers across multiple subjects by defining standards for acceptable patterns of strengths and weaknesses across subjects. While elementary teachers are highly unlikely to be fully expert in all subjects, public policy considerations suggest that serious weaknesses, particularly in subject areas deemed to be most fundamental (e.g., reading, mathematics), should be detected and remedied. How to represent this policy concern in evaluation methodology for elementary teachers remains problematic.

One might expect that the CNTP innovative assessments and the other assessments might be complementary, with each shedding light on different teaching domains. However, given the high number of contradictory assessment results among the relatively small number of teachers interviewed, this seems to be too simplistic a description. Moreover, in the majority of cases where evaluations differed, enough assessment information was provided to suggest that each assessment decision was basically sound, based on the evidence gathered. This suggests, as has been the CNTP's contention all along, that conceptual development of a systematic approach to assessment is necessary before any innovative assessments are considered for implementation. A valid and reliable methodology of sampling across occasions and topics must also be developed, and the ability of alternative assessment results to predict the application of teaching skills must be strengthened. However, the cases of Grace, and perhaps Deanna, also suggest that district

assessments do not always identify teacher weaknesses, suggesting a possible role for improved alternative assessments.